

IntelliGenetics

FastDB - Fast Pairwise Comparison of Sequences

Release 5.4

Results file .seg2-4-6-8.res made by mruhl on wed 18 Jun 103 15:20:25-PDT

Query sequence being compared:	US-10-081-644-2	(1-370)
Number of sequences searched:		6
Number of scores above cutoff:		6

Results of the initial comparison of US-10-081-644-2 (1-370) with
File : US10081644.pep

Normal distribution curve showing the distribution of scores. The mean is 123, and the standard deviation is 41. The x-axis is labeled 'STDDEV' and the y-axis is labeled 'SCORE'.

Score	Standard Deviation (STDDEV)
0	0
41	41
82	82
123	123
164	164
206	206
247	247
288	288
329	329
370	370

PARAMETERS		
Similarity matrix	PAM-150	2
Threshold level of sim.	16%	
Match penalty	5.00	Joining penalty
Gap penalty	0.05	Window size
Gap size penalty	0.05	
Cutoff score	1	
Randomization group	0	

SEARCH STATISTICS

Scores:	Mean	Median	Standard Deviation
---------	------	--------	--------------------

```

Times:      CPU      Total Elapsed
00:00:00.00 00:00:00.00

```

```

Number of residues:      1515
Number of sequences searched: 6
Number of scores above cutoff: 6

```

The scores below are sorted by initial score. Significance is calculated based on initial score

A 100% identical sequence to the query sequence was found:

Sequence Name	Description	Init.	Opt.	Length	Score	Score	Sig.	Frame
---------------	-------------	-------	------	--------	-------	-------	------	-------

1. US-10-081-644-2 Sequence 2, Application US 370 370 1.29 0

Sequence Name	Description	Length	Score	Int. Opt. Score	Sig.	Frame
2. US-10-081-644-4	Sequence 4	368	245	291	0.42	0
3. US-10-081-644-8	Sequence 8	376	237	283	0.36	0
4. US-10-081-644-6	Sequence 6	377	237	287	0.36	0

1. US-10-081-644-2 (1-370)
US-10-081-644-2 Sequence 2, Application US/10081644

Initial Score	=	370	Optimized Score	=	370	Significance	=	1.23
Residue Identity	=	100%	Matches	=	370	Mismatches	=	0
Gaps	=	0	Conservative Substitutions	=	0			

MSVPTTOKAVIIIEBGRKAVVKTVDSVPELKEGTALVYVEAAGNPTDMKIIAYKIGBESSIIIGCDIAGTVVKI
X 10 20 30 40 50 60 70
MSVPTTOKAVIIIEBGRKAVVKTVDSVPELKEGTALVYVEAAGNPTDMKIIAYKIGBESSIIIGCDIAGTVVKI
MSVPTTOKAVIIIEBGRKAVVKTVDSVPELKEGTALVYVEAAGNPTDMKIIAYKIGBESSIIIGCDIAGTVVKI
X 10 20 30 40 50 60 70
GNASTDLKAVDGTGFGFPHGASQTDPNNGAFAEYARVYRPLFYKSNLTSTDAEISEGCVKNFESAASLIPVS
80 90 100 110 120 130 140
GNASTDLKAVDGTGFGFPHGASQTDPNNGAFAEYARVYRPLFYKSNLTSTDAEISEGCVKNFESAASLIPVS
80 90 100 110 120 130 140

220 230 240 250 260 270 280
 DVEFPIHAGVTEQKSKYPLQVINDAVSBSIIPAYKYTDSIPATLLEVPMTTISFEERKDNVK
 DVEFPIHAGVTEQKSKYPLQVINDAVSBSIIPAYKYTDSIPATLLEVPMTTISFEERKDNVK
 220 230 240 250 260 270 280
 LTTAAVSLCHLIGSMEMHPSTPHTPLILMGCAVVGQQLIQVAKHINATKYITVASKKHGLKMSYSG
 LTTAAVSLCHLIGSMEMHPSTPHTPLILMGCAVVGQQLIQVAKHINATKYITVASKKHGLKMSYSG
 150 160 170 180 190 200 210

[illegible]

2. US-10-081-644-2 (1-370)

Initial Score	=	245	Optimized Score	=	291	Significance	=	0.42
Residue Identity	=	53%	Matches	=	201	Mismatches	=	136
Gaps	=	3	Conservative Substitutions	=			=	33

[illegible]

KLGPNDAAFAIGDIYGVIGHGASVFPFSGNAGFAEYSAISSETAYKPAPEFRLOGDKLPEGEVKSLEGA
 80 90 100 110 120 130 140

140 150 160 170 180 190 200 210
 SLVPSLTAGVSLCHHLGSKMEHPSTPQHTPLLIWGATAVGQOLIQAACHINATKYIVASKGHEKLL
 150 160 170 180 190 200 210
 SLVPSLTAGVSLCHHLGSKMEHPSTPQHTPLLIWGATAVGQOLIQAACHINATKYIVASKGHEKLL

220 230 240 250 260 270 280
 KSYGADVDFDYHDAGVTEQIKSKYPNLOHVIDAVGSEDSIPEAYKTADSLPATLLEVPMTIESIPEIRK
 220 230 240 250 260 270 280
 KSYGADVDFDYHDAGVTEQIKSKYPNLOHVIDAVGSEDSIPEAYKTADSLPATLLEVPMTIESIPEIRK

290 300 310 320 330 340 350
 DNKIDITLLYRASGOEILLGATRFPSPEYHEATVVKFVKNPHLNGDIHMKIKVPSNGLDVVALTEG
 290 300 310 320 330 340 350
 DNKIDITLLYRASGOEILLGATRFPSPEYHEATVVKFVKNPHLNGDIHMKIKVPSNGLDVVALTEG

360 370
 IKGKKNKVKYVARL
 370 X
 IKHNGSGEKLVAVLN

US-10-081-644-2 (1-370)
 Application US/10081644

Initial Score = 237 Optimized Score = 283 Significance = 0.36
 Residue Identity = 52% Matches = 194 Mismatches = 143
 Gaps = 3 Conservative Substitutions = 27

X 10 20 30 40 50 60 70
 MSVPTTKAVYIIEGDKAVVKTIVSVPELKEGTALVKVEAVAGNPTDKHIAKYKIGPESILGCDIAGTVL
 10 20 30 40 50 60 70
 MSVPTTKAVYIIEGDKAVVKTIVSVPELKEGTALVKVEAVAGNPTDKHIAKYKIGPESILGCDIAGTVL

80 90 100 110 120 130 140
 GPNA-STDLKVGDTGFGFVHGAQTDPRKNGAFAYARVYPLFYKS--NLTHSTADEISBGPVKNPESASL
 80 90 100 110 120 130 140
 GPNA-STDLKVGDTGFGFVHGAQTDPRKNGAFAYARVYPLFYKS--NLTHSTADEISBGPVKNPESASL

150 160 170 180 190 200 210
 PVSLLTNGVSLCHHLGSKMEHPSTPQHTPLLIWGATAVGQOLIQAACHINATKYIVASKGHEKLLKS
 150 160 170 180 190 200 210
 PVSLLTNGVSLCHHLGSKMEHPSTPQHTPLLIWGATAVGQOLIQAACHINATKYIVASKGHEKLLKS

220 230 240 250 260 270 280
 YGADVDFDYHDAGVTEQIKSKYPNLOHVIDAVGSEDSIPEAYKTADSLPATLLEVPMTIESIPEIRKDN
 220 230 240 250 260 270 280
 YGADVDFDYHDAGVTEQIKSKYPNLOHVIDAVGSEDSIPEAYKTADSLPATLLEVPMTIESIPEIRKDN

290 300 310 320 330 340 350
 VKIDITLLYRASGOEILLGATRFPSPEYHEATVVKFVKNPHLNGDIHMKIKVPSNGLDVVALTEGK
 290 300 310 320 330 340 350
 VKIDITLLYRASGOEILLGATRFPSPEYHEATVVKFVKNPHLNGDIHMKIKVPSNGLDVVALTEGK

360 370
 ESKKKNKVKYVARL
 370 X
 IKGKKNKVKYVARL

US-10-081-644-6 Sequence 6, Application US/10081644

Initial Score = 237 Optimized Score = 287 Significance = 0.36
 Residue Identity = 51% Matches = 193 Mismatches = 145
 Gaps = 3 Conservative Substitutions = 32

X 10 20 30 40 50 60
 MSVPTTKAVYIIEGDKAVVKTIVSVPELKEGTALVKVEAVAGNPTDKHIAKYKIGPESILGCDIAGTV
 10 20 30 40 50 60
 MSVPTTKAVYIIEGDKAVVKTIVSVPELKEGTALVKVEAVAGNPTDKHIAKYKIGPESILGCDIAGTV

70 80 90 100 110 120 130 140
 VKLGPNA-STDLKVGDTGFGFVHGAQTDPRKNGAFAYARVYPLFYKS--NLTHSTADEISBGPVKNPESA
 70 80 90 100 110 120 130 140
 VKLGPNA-STDLKVGDTGFGFVHGAQTDPRKNGAFAYARVYPLFYKS--NLTHSTADEISBGPVKNPESA

150 160 170 180 190 200 210
 ASLPSLTAGVSLCHHLGSKMEHPSTPQHTPLLIWGATAVGQOLIQAACHINATKYIVASKGHEKLL
 150 160 170 180 190 200 210
 ASLPSLTAGVSLCHHLGSKMEHPSTPQHTPLLIWGATAVGQOLIQAACHINATKYIVASKGHEKLL

220 230 240 250 260 270 280
 LKSYGADVDFDYHDAGVTEQIKSKYPNLOHVIDAVGSEDSIPEAYKTADSLPATLLEVPMTIESIPEIR
 220 230 240 250 260 270 280
 LKSYGADVDFDYHDAGVTEQIKSKYPNLOHVIDAVGSEDSIPEAYKTADSLPATLLEVPMTIESIPEIR

290 300 310 320 330 340 350
 KDNKIDITLLYRASGOEILLGATRFPSPEYHEATVVKFVKNPHLNGDIHMKIKVPSNGLDVVALTE
 290 300 310 320 330 340 350
 KDNKIDITLLYRASGOEILLGATRFPSPEYHEATVVKFVKNPHLNGDIHMKIKVPSNGLDVVALTE

360 370
 GIKGKKNKVKYVARL
 370 X
 DIKKNKSGEKLVAVLN

```
620      630      640      650      660      670      680
GATGGGTGCCATTCCATTTTGGAGCCCAAGATGACACAAACTTACACCGAGTGTCAAC---GAACT
|||
AATTGTGAAGAGTACGAGTGAAGTAACTTTTGACTACAGAGTGTAGCTTATCGAACAGATAAAA
650      660      670      680      690      700      710
690      700      710      720      730      740      750      760
GGCAATATGTCACAGATTTGGAATTTTTCACAGAGGCTTCAGAAATTTTCATCAGCAGTGTGAGTTAG
|||
AGAGATACAAACAATTCCTTACTGTGTGAGTGTGTCTCAACACAGAAACTATTCACAGAGTACAAAT
720      730      740      750      760      770      780
770      780      790      800      810      820      830
TTACTCTGTAAACAAAGGTGATTAACCTTGCAATATTCAGAAATGCAACCATTTTGTAGATCTGTGG
|||
GTGCCCCGTGATGACTTAGACGCTACGGTCTC---GTTCAATTGACCGTTTAAACGAAAAGATATCAAGAGG
800      810      820      830      840      850
840      850      860      870      880      890      900
GAAGCACCGGTGAACAAACCGAAACCGGTATCTCCAACTTCAAGTCACTAGACTAGCTTTGCTCAAGTTG
|||
AAGACAGAGGCAAAACGTAGTATTTGAAGAAACCTTCTATATTTGATAGAGATACGACGTCCTCATTTG
860      870      880      890      900      910      920      930
910      920      930      940      950      960      970
ACAACTGTACCAAGATGTCAATCTTAGAATTGAACCTTCTGAGCAATCTTATATAGCAATATGCTTCAA
|||
GCACGTTTACTTTGCCAGACACCTGATTAACAAGAGAGCGCCATATAATTATTTAGTTCAATCAATCCA
940      950      960      970      980      990      1000
980      990      1000      1010      1020      1030      1040
TCAGTTGGTTACGACAGACAGCTTCACTTCAATGATGACGGCTTCTTAATTTGGAATGAGACA
|||
AAATCAATGATGTAATCCACCACTCCAGTGAAGTTTACAGAAACGGTTGATGATATCCACAGCT
1010      1020      1030      1040      1050      1060
1050      1060      1070      1080      1090      1100      1110      1120      1130      1140
TCGTTTATACACAGCTTGTCTACCTTCAATGATGACGGCTTTGAGTGTGAACTGACAT
|||
TACTTGATGATTTTACGACGAGGAAATTCGGCGAAAGTTGCTTCCGCTTGTGAATATATCTAGACTG
1080      1090      1100      1110      1120      1130      1140
```

[illegible]

```

CAAGGCTATATGAAGACGGGGCTTTACAGATGTTCTCGTATCTCGAAGACATTAAATCGTAAAGAACTCTG
1020      1030      1040      1050      1060      1070      1080      1090
GTGACAAACCGGAAACCGGATATCTCCAAACCTTCATAGTACTAGACATTTGGTCCAAAGTTGACAACTGT
1100      1110      1120      1130
GTGAAAACTCGTGGCCGGATTTAAACTAATCTTAAGAC
1140      1150      1160      1170      1180      1190      1200
920
ACGACGAATGTCACAT

4. US-10-081-644-1' (1-1113)
US-10-081-644-3 Sequence 3, Application US/10081644 /

Initial Score = 42 Optimized Score = 359 Significance = 1.66
Residue Identity = 36% Matches = 414 Mismatches = 693
Gaps = 28 Conservative Substitutions = 0

X
10 20 30 40 50 60
TTATACCTCGCAACATCTTAACTATTTGGTTTAACTCTTTATTAACCTTCAGTAGAG
10 20 30 40 50 60
CAAAATGCTCGCTCGATTCACAAACCAATGAAGACCGGTTCATGAAATGGCAAGCTGTAGTCAAAAC
10 20 30 40 50 60 70
CTGGGACATC-ATCTAACCGCGTGGCTGTAACCTTAAATATTCATATGATGATACACCGTGTTAAGGTGT
10 20 30 40 50 60 70 80 90 100 110 120 130 140
AGGACATTCACAAATCTGATTTAGAGAGAGATTTGTTCTTAATTAAGACGTGCGCGTGGCCGATACCTTA
10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170 180 190 200
GGAATTATTAACCTTACCAATTTAACTATGCTGGCTTACGATATTTCTGACATGACAGGAAATCTTGTGGACCC
10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170 180 190 200
CCGATTGAAACATATGATTTCAAGATTGG-TCTCAAGTGCCTCTTAAGCGTGAATGACGGCCAA
10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170 180 190 200 210
AATAGATTTCTTGGCA-----GATGCAAGATACAA-AGATATATCAATTTTAACTATCTTT
10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170 180 190 200 210
ATGTAAGTTGGGCAAAATGTGATGCTGACAGCTTGGCCATGTGATTTACATTATGGGTTATTAC
10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170 180 190 200 210
TCGATTTCTCAGGAATGCTTCAATGTCATGTAGAACCACTTCAATTAATGAGGAGTAGACTCT---
10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170 180 190 200 210
GGTGTTCAGTAGAGTTCCCTCAACGGTGGCTTGGTAGACTCTCCATTTCAATCCGAGCTGCTTAT
10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170 180 190 200 210
340 350 360 370 380 390 400
---GCTGACATTATPAGCCCTCGGGATATCTATCTTGGCTTCCCAAGCGTCAATTAACATGTTCCAGTTT
10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170 180 190 200 210
AAACCAAGCCAGAGATTTAGATTTGGGGTAAAGACAGCTACCAAGGCCCGTAATCTTTAGAGAGG
10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170 180 190 200 210
410 420 430 440 450 460 470
GGAATTCGATTTGATCTGCTCATATACCGCTGCATCATGATATGCAAGACATCATCAG-ACCATTAAGA
10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170 180 190 200 210
GCAGTATCCCTCCAGGTCTTATGACCAAGGCTG--GTATGATCTTAAACATATGTTGGTTGGAATGA
10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170 180 190 200 210
440 450 460 470 480 490 500
CATGAAAGCCCTCAAAACGCAAGAAATCAACCACTTATTTTGGGGTGGTGCACATGCTGTGGCCAGAA
10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170 180 190 200 210
510 520 530 540 550 560 570
CTTAAAGCTTTTATGCTTTTATGAGACAAAGCTTACATCTTATGATTTAAGCATATGATTTTGGCAAC
10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170 180 190 200 210
540 550 560 570 580 590 600 610
TTGATTAATGTTTATGACCACTGCTGTAGACCAACCCCAATCAATATGATGATGATGTT--GCGGGGTAT
10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170 180 190 200 210
610 620 630 640
TGCTTTTATATGGCAAAAATTAAACGTTTACAGAGATCATCTGCTGCTCTCTGTAACATGAAAA
10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170 180 190 200 210
580 590 600 610 620 630 640

```



```

--CAAGTATGACAACTACACAGCAGTTGTCAACGAACTGGCAATGACAGCATTCGAAGTTC
630 640 650 660 670 680 690
GAAAGCTTTAAAGCTTATGAGTGGATGATGTGTACATCATGACGAGGCTATTGAGAGATC
700 710 720 730 740 750 760
TTACAGGGGCTTCAGAAATTTTATCATGACAGTTGA--GTGAGTTAACTCTCTTGAACAAAGTGTGAT
770 780 790 800 810 820 830
AAATCGAAGTATCCAAACCTGCAACATGTTATTGACGCTGGGAGCGAAGATGATCCCGAGGCTAT
840 850 860 870 880 890 900
790 800 810 820 830 840
AAA-CCCTGGCATATTCAGCAAAATGCAACATTTTATGATCTGT-----TTGGAGACACCGTGAA
850 860 870 880 890 900 910
GAAATCAGAAAGATATGTTAAATGATTAATCTTGTGTATGTCGATCGTGTCAAGAAATCTATTG
840 850 860 870 880 890 900
920 930 940 950 960 970 980
GTACAGCAATGTCAATCTAGAAATTTGAACCTTCTGGA--CCAATCTTATGCAATATCTTCCAAATCAGT
990 1000 1010 1020 1030 1040 1050
GGTGCAACAGATTTCTGCTAGTCCAGCAATATCATGACAGCAAGTTAAATGTTAATTAATCA
920 930 940 950 960 970 980
990 1000 1010 1020 1030 1040 1050
TGGATTACACGCAACACCTTCACCTTCCAGGCTGTACCTCTTAAATCTGGAACGTGACATCTGT
1060 1070 1080 1090 1100 1110 X
TTTAAACAAGCTTTGTACCTTCAATGATGACGGCTTTTGAAGTGTGGAATGTCAT
1060 1070 1080 1090 1100 1110 X
TTCACCTGAAAGTATTAAGAGTAAACAAATGTTAATGATGTTGCGCATGATTA
1060 1070 1080 1090 1100 1110 X
2. US-10-081-644-1' (1-1113)
US-10-081-644-5 Sequence 5, Application US/10081644
Initial Score = 45 Optimized Score = 300 Significance = 1.87
Residue Identity = 36% Matches = 332 Mismatches = 567
Conservative Substitutions = 0
X
CTCAAGATCTATCTGGAGATGATGCTGTGCGCAATTTGCAATTTGGCGCCAGCTGTCAATCTTAAG
180 190 200 210 220 X 230 240
30 40 50 60 70 80 90
ACATTTTGTGTTTACCTCTTTTATATCACTTCACTGAGAGCTGGACATCATCTTAACCTGTAAGT
250 260 270 280 290 300 310
ACTTTTCTATCGGATTAATTTATATGAGGCTTCAACGATCTTCGTAAGTTCCTTCCATGAGTCT
100 110 120 130 140 150 160
TTAATATTCATATATGATGATATCACCTGTTTAAAGGTGATTAATACTTAACGAATTAATTAATCT
170 180 190 200 210 220 230
TCATGATATTCATGATGCAAGAAATCTGTGTACCCCAATAG--AATTTCTGACCAAGATGACGATACA
320 330 340 350 360 370 380
GGAATGTTCTATCTGCGGCGCTGTCAAGTCTTTGGAAGGTGACACTATCCAGTGTCACTGACCAAC
390 400 410 420 430 440 450 460

```

```

240 250 260 270 280 290 300
ACAAATATATATCAATTTTAACTATATCTTTCTGTATTTCTTACAGAAATGCTTCAATGATGGAACCA
470 480 490 500 510 520 530
CGCTTGTTGTTGACCTTAATATCTTGGGCTTGGACCTGAGTGGAGCCATCAACCCACAAAGAAAGGCC
540 550 560 570 580 590 600
310 320 330 340 350 360 370 380
CTTCTAATATATGTCAGGATAGACTATATCTGTGTGACTTTATAGGCTCGGGATATATCTTGCTTCCA
540 550 560 570 580 590 600
CATCTTATATATGCGGCGGTCACATGCAATAGGTGAGTGTGCTCATCTCAATTTGCAATTAATGATGCTT
540 550 560 570 580 590 600
390 400 410 420 430 440 450
CAGCTGCAATATACATGTTGTGAGTTGATCTTCAATCTGCTGATTAATGACCCCTGATCATGATAGT
610 620 630 640 650 660 670
CACTCAATATCATTTGTGTGTGCTTCTCGAAGACGAAAACTTTTC--AAAGATATGAGTGTGCTGATTA
460 470 480 490 500 510 520
CAAAAGCATCATGACGACCATTAAGACTTTTAAAGCTTTTCAATGCTTTTATGAGCAACAGTTACAAATCTTAG
680 690 700 710 720 730 740
TTTGATTTATCATGATATTTGACGCTGTGAAACAAATTAACAAGATTAACAATATCTGTATTTATGTCGAC
530 540 550 560 570 580 590
TATACCATATGATATGTTTGGCAACTGTGATTTAGTTGATGACCTGCTGTGACCAACCCCAATCAATA
750 760 770 780 790 800 810
TGTGTGGCAATCATGATATGCTTCAACAAAGTGTACAAATGTGCGGCGATTAACAGATGCTACAAATGTT
600 610 620 630 640 650 660
ATGATGATGATGTTGCGGGGTATGATGAGTGGTCCATTCATTTTATGAGCCCAAGTATGACAACTAAAC
820 830 840 850 860 870 880 890
GAATTTAAATTTTACAGAAAGAAACGCAAAAAAGAAACAGAGCAAAACGTTACTATGATGATTAATA
670 680 690 700 710 720 730
CAGCAGTTGTCAACGAACCTGGAGATGACAGCAGATTCGAATGTT---CTTACAGGCGCTTCAGAAATTT
900 910 920 930 940 950 960
AGGCTATATTCAA--TAGGTGCGCATGAGATGACCAATTTGGAACATTTACTTTTCCAGCGGACACAGAGT
740 750 760 770 780 790 800
CATCAGCAGTTGAGTGAATTAATCTC-----TTGTAAACAAAGGTGATTAACCTTGGCATATTCAGC
970 980 990 1000 1010 1020 1030
GGAAGCTGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT
810 820 830 840 850 860 870
AAATGCAACATTTTATGATCTGTTTGGAGAGCAGCTGTGAAACAAACCGAAACCGGATCTCCACCTTCAA
1040 1050 1060 1070 1080 1090 1100
TAAGGCTTAAAGAACGGGCTTGTGATGTTCTCTCATATCTTAAGACATCAATATGTAAGAACTGTG
880 890 900 X 910 920 930 940
GTACAGTATGACATTTGTGTCACAGTTTGAACACTGTACAGCAATGTCACATCTTAATGAACTTCTG
1110 1120 1130 X
GTAAGAAACCTGTTGCGGTATTAACATA
950
ACCAATC
3. US-10-081-644-1' (1-1113)
US-10-081-644-7 Sequence 7, Application US/10081644
Initial Score = 44 Optimized Score = 296 Significance = 1.80
Residue Identity = 38% Matches = 359 Mismatches = 513
Conservative Substitutions = 0
X
TTTAAACCTGCAACATACTTA

```

IntelliGenetics

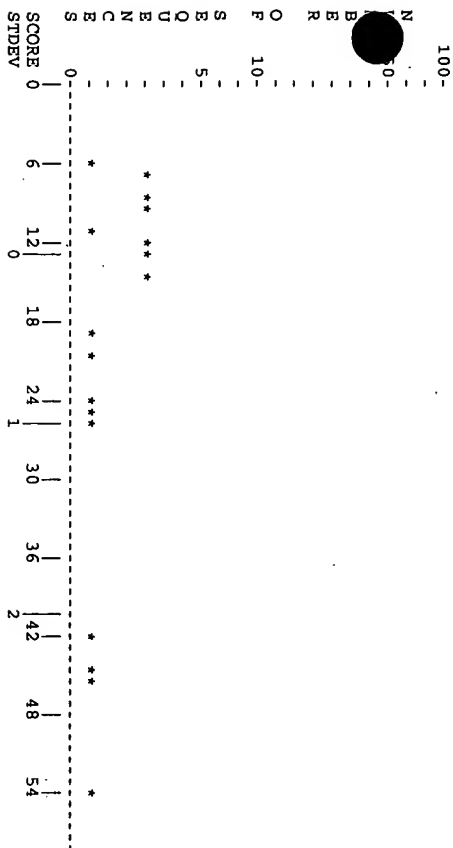
FastDB - Fast Pairwise Comparison of Sequences

Release 5.4

Results file beginlv-3-5-7.res made by mruhl on wed 18 Jun 103 15:30:48-PDT.

Query sequence being compared:US-10-081-644-1'	(1-11113)
Number of sequences searched:	23
Number of scores above cutoff:	23

Results of the initial comparison of US-10-081-644-1' (1-11113) with:
File : US10081644.seq



	Unary	K-tuple
Similarity matrix	1	4
Mismatch penalty	1	300
Gap penalty	5.00	Window size
Size penalty	0.33	500
Self score	1	
Randomization group	0	

SEARCH STATISTICS

Scores:	Mean	Median	Standard Deviation
	19	14	13.90
Times:	CPU		Total Elapsed
	00:00:00.00		00:00:00.00

Number of residues:	5519
Number of sequences searched:	23
Number of scores above cutoff:	23

The scores below are sorted by initial score.
Significance is calculated based on initial score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Length	Init. Score	Opt. Frame	Sig.

1. US-10-081-644-1	Sequence 1, Application US/10081644	1113	54	378	2.52
2. US-10-081-644-5	Sequence 5, Application US/1134	45	300	1.87	0
3. US-10-081-644-7	Sequence 7, Application US/1122	44	296	1.80	0
4. US-10-081-644-3	Sequence 3, Application US/1145	42	359	1.66	0

**** 2 standard deviations above mean ****					
Initial Score	= 54	Optimized Score	= 378	Significance	= 2.52
Residue Identity	= 48	Matches	= 460	Mismatches	= 625
Gaps	= 54	Conservative Substitutions	= 0		

1. US-10-081-644-1' (1-1113)					
US-10-081-644-1 Sequence 1, Application US/10081644					

X	10	20	30	40	50
TTTAACCTGGCAACATCTTAACATTTTGTTTTACCTTTTATACCTTCAGNAGAGCTGGACATC					
ATGTCAGTTCACCAACCTAATAAAGCCGTCATCATGAAGGACAAACCTGTTGTTAAACAGATGCTC					
X	10	20	30	40	50
ATCTAACCCGTCGTGTAACCTTTAATATTCAT	ANGATGAGATATCCGTTGTAAGGTGATTTATTA				
AGTTCAGAAATTAAGAGGGTACAGCCTTGTTGTAAGGTGAGGCGTGTCTGGTATCCCACTGATGATGA					
X	80	90	100	110	120
150	160	170	180	190	200
ACTTAACGAATTTAACTGTCGTGCTTCATGATATTTCTGACATGACAGAAATCTTGTCACCACTAATAATTT					
GCAATATGCTTAAATGAAGTTG3-TCCAAAGAGTTCAATTTAGAGATGCAATGCTGTAACAGTTGTCAAAC					
X	150	160	170	180	190
220	230	240	250	260	270
CTTGACCAAGATGACCGCATTAACAAGATATATCAATTTTAACATTTCTTTTCTGATTTCTTCAGAAATGC					
TTGGACCAAAATG-----CTAGTACTGACTGTAAGGTGTAAGGTACCGGTTTGTGTTGTCACGAGTCT					
X	220	230	240	250	260
280	290	300	310	320	330
TTTAAATGTCATTTGGAACCACTTTATATATGTGGCAGGTAGACTATCTGCTGTGACTTTATATAGGCTCGG					
TCCCA-----AACGATTCCTAAATATGATGATTTGCTGATATATGC-CAGGATTTATCCACCTTTG					
X	280	290	300	310	320
360	370	380	390	400	410
GGATACTATCTTTGCTTCCACACAGCGTCATTAACAATGTTSCAGGTTTGATTAATCTTGATTTGATCTGCTCAA					
TTTTACAGAAGTAACTTAATCTCA--CTCAACTGCTGATGAATTTTGGAAGGCCCTGTGAAGAATTCGAAT					
X	340	350	360	370	380
440	450	460	470	480	490
TAAACCGTCATCATGATGTATGTCAAAGCATCATCAGACCATTAAGATTTAAAGCTTTTCATGCTTTTATG					
CTGTGTCATCATGTCCAGTTTCTGTGCAACTCTGTG---GTGTTAGTTTGTGTCAATCACTTGGGCTCAAAA					
X	410	420	430	440	450
510	520	530	540	550	560
AAGCAACAGTTAACAATTTAGTATTAACCAATGTAATGTGTTGGCAACTGTGATTAAGTTAGTGGACCACTGCTG					
ATGGAATGAGCAACCATCTCAACCCGCAACATACATCATCATTAATTTGATTTGGGGTGGGTGCTACAGAGTGAGT					
X	480	490	500	510	520
580	590	600	610	620	630
TAGACCAACCCCAAAATCAATAATTAATGATGATGATGTGGCGGAGATGAGTGGTGCATTTCAATTTTGAGCC--					
CAACCAATTAATCCAAAGTGGCAACATATCAATGCTTAATCTAAGATTTGAATGTTGCTTTAAAAAAGCAT					
X	550	560	570	580	590
650	660	670	680	690	700
710					

AAAGAGTAAACAAAATGTTAGATGTTCCAGCTTATA
 |||||
 AAATCGGTAAGAACTGTGTAAGAACTCGTTCCTATTAAACTATCTAGAAAC
 |||||
 1070 1080 1090 1100 1110 1120

10 X 20 30 40 50 60 70

70 80 90 100 110 120 130

AGATGTCTCAGTTCAGAAATTTAAAGAGGGTACAGCGTTGGTGAAGGTGAGCGTGTGTGTGTAACCCAA

AGACATCTCAATCTCAATTGAAATTGAAAGAAAGATTGTTGTTCTTAATTAGACTGTGCGCGTTGCCGTAACCTTAA

80 90 100 110 120 130 140

CTGATTGAGACATATGTGCTTAAAGATGTGCTCAGAGGTTCAATTTCTAGAGTGTGACATTGTCTGTACAG

CCGATTGGAAACAATTTGATTTCAAGATTGTCTCTCAAGGTGCCCTTAAAGCTGTGATGACAGCGGCCAA

150 160 170 180 190 200 210

TTGTCAAACTTGGACCAATGTCTGACTG---ACTTGAAGTGTGAGATACCGGTTTCGTTTGTTCACG

TGCTAAAGTTGGGCCCAAAATGTTATGCTGACGCTTTCGATTGTGTGATTACATTATAGGGCTATTACG

230 240 250 260 270 280 290

GTGTCTCCCAACAGATCTTAAATATGTGTCATTGTGTGAAATATGTCAGAGGTTTATCCACCTTGTGTTTAC

GTGCTTCAGTGAAGGTTCCCGCCAAAGCGTCCCTTGTGTGATCTGTGCATTATATCCGACATGCTTATAT

290 300 310 320 330 340 350 360

A-----GAGTAACTTAACTCACTCACTCACTGATGAAATTTCTGAAGGCCCTGTGAAGAACTTGAATCTGA

AACCGACAGAGAGTTTGAATTGTGGCGGTAAAGCAAGCTACAGAGGCCCGTAATCTTTGAGAGGGG

370 380 390 400 410 420 430

CTGCATCACTTGCAGAGTTCTGTGTAACAACCTGCTGATGTGTGTGTCTATCACTTGTGGCTCAAAAATGAAAT

CAGTATCCCTCCAGTCTCATATGACACGCGTGTATGATCCTTACACTAATTTTGTGGCTTGACATGACAA

440 450 460 470 480 490 500

490 500 510 520 530 540 550

GGCACCCACTTACCCCGCAACATACCTCATCTATATGATTTGGGGTGTGTCTACAGACAGTGGGTCAACAA

GGAACCCCTCCAAAGCCGGAAGATCAACCATCTTATTTGGGGTGTGTGCACGCGTGTGGCAATGCG

510 520 530 540 550 560 570

560 570 580 590 600 610 620

TATATCAAGTTGCCAAACATATCATCTTATCTAAGATTGTAATCTGTCTGTCTTAAAAAGCATGAAGAAG

TTATTCATTTGGCAAAAACCTAAACGTTTGACGAAGTCAATCGTCTGTCTTCTCGTAAACATGAAGAAAT

580 590 600 610 620 630 640

630 640 650 660 670 680 690 700

TTTTAAAGCTTTATGTGTGCTGATGATGTCTTTTACATCATGATGACAGCGCTTATTTAGACAGATCAATCGA

GTGTTGAAGAGTAAAGGTGCGAGATGAATTTTGTGACACAGATGTGACGTTATCGAACAGATTAAGAAAG

650 660 670 680 690 700 710 720

710 720 730 740 750 760 770 780

AGTATCCAAACCTGCAACATGTTATTAACCGCTGTGGAAAGCGAAGATAGTATCCCGGAGCGCTTAAAGTCA

AGTCAAAACATTTCTTACTGTGTGATCTGTGTCTTCCACACAGAAATTAATTCAAAGGTGTACAAATGTG

730 740 750 760 770 780 790 800

800 810 820 830 840 850 860 870

CAGCAGATGATCTACCTGCGACACATATTTAAGGTGTTCACATGACATTTGAAGACATTTCTTGAAGAAATCA

CCCGTGAATGATTTAGAGCTACGCGTCTTCAATTAAGACCGTTTAAACCAAAAAGATTCAGAGAGGAAGACA

880 890 900 910 920 930 940 950

GAGGCGCAAAAGCTCAGATTTGAAGAGAACCTTTATATTTGATTAAGAGGTAAACAGATCCCATTTTGGACGT

870 880 890 900 910 920 930

330 340 350 360 370 380 390
TTTATCACCTTGGTTTACAAGAGTAATCTACACTCAACGTCGATGAATTTTGAAGGCCGTGTA
|||||

AGTAATTACCTACTCA-----CTGCGATGAATAATTCTAAGCCCTGTGAGAATTGGAACTC
AATCACCATGAATCTCAAATTTTGGGAGAGACTTCTACTCCGCCCTGTACAGTCCTTGGAAAGTC
360 370 380 390 400 410 420 430

> O < IntellGenetics
> O <

FastDB - Fast Pairwise Comparison of Sequences
Release 5.4

Results file seq1-3-5-7.res made by mrhl on Wed 18 Jun 103 15:27:36-PDT.

Query sequence being compared: US-10-081-644-1 (1-1113)
Number of sequences searched: 23
Number of scores above cutoff: 23

Results of the initial comparison of US-10-081-644-1 (1-1113) with:
File : US10081644.seq

```

100-
N -
S -
E -
O -
F -
R -
E -
O -
S -
E -
U -
N -
C -
E -
S -
SCORE 0 124 247 371 495 618 742 866 989 1113
STDEV

```

PARAMETERS

Similarity matrix Unitary 4
Mismatch penalty 1 Joining penalty 30
Gap penalty 5.00 Window size 500
Gap size penalty 0.33
Self score 1
Randomization group 0

SEARCH STATISTICS

Scores: Mean 135 Median 14 Standard Deviation 271.04

Times: CPU 00:00:00.00 Total Elapsed 00:00:00.00

Number of residues: 5519

Number of sequences searched: 23

Number of scores above cutoff: 23

The scores below are sorted by initial score.

Significance is calculated based on initial score.

A 100% identical sequence to the query sequence was found:

Sequence Name	Description	Init. Opt.	Length	Score	Score	Sig.	Frame

1. US-10-081-644-1 Sequence 1, Application US 1113 1113 1113 3.61 0
The list of other best scores is:

Sequence Name	Description	Length	Score	Sig.	Frame		
**** 1 standard deviation above mean ****							
2. US-10-081-644-14 Sequence 14, Application US		509	504	1.36	0		
3. US-10-081-644-5 Sequence 5, Application US		1134	427	1.08	0		
4. US-10-081-644-3 Sequence 3, Application US		1145	427	1.08	0		
5. US-10-081-644-7 Sequence 7, Application US		1122	422	635	1.06		
1. US-10-081-644-1 (1-1113) US-10-081-644-1 Sequence 1, Application US/10081644							
Initial Score = 1113	Optimized Score = 1113	Significance = 3.61					
Residue Identity = 100%	Matches = 1113	Mismatches = 0					
Gaps = 0	Conservative Substitutions = 0						
X	10	20	30	40	50	60	70
ATGTCAGTTCACCACTTAATAAGCCGTCATCATTTGAAGGTGACCAAGCTGTTTAATAAGATGTCTCA							
ATGTCAGTTCACCACTTAATAAGCCGTCATCATTTGAAGGTGACCAAGCTGTTTAATAAGATGTCTCA							
X	10	20	30	40	50	60	70
ATGTCAGTTCACCACTTAATAAGCCGTCATCATTTGAAGGTGACCAAGCTGTTTAATAAGATGTCTCA							
ATGTCAGTTCACCACTTAATAAGCCGTCATCATTTGAAGGTGACCAAGCTGTTTAATAAGATGTCTCA							
X	10	20	30	40	50	60	70
ATGTCAGTTCACCACTTAATAAGCCGTCATCATTTGAAGGTGACCAAGCTGTTTAATAAGATGTCTCA							
ATGTCAGTTCACCACTTAATAAGCCGTCATCATTTGAAGGTGACCAAGCTGTTTAATAAGATGTCTCA							
X	10	20	30	40	50	60	70
ATGTCAGTTCACCACTTAATAAGCCGTCATCATTTGAAGGTGACCAAGCTGTTTAATAAGATGTCTCA							
ATGTCAGTTCACCACTTAATAAGCCGTCATCATTTGAAGGTGACCAAGCTGTTTAATAAGATGTCTCA							
X	10	20	30	40	50	60	70
ATGTCAGTTCACCACTTAATAAGCCGTCATCATTTGAAGGTGACCAAGCTGTTTAATAAGATGTCTCA							
ATGTCAGTTCACCACTTAATAAGCCGTCATCATTTGAAGGTGACCAAGCTGTTTAATAAGATGTCTCA							
X	10	20	30	40	50	60	70
ATGTCAGTTCACCACTTAATAAGCCGTCATCATTTGAAGGTGACCAAGCTGTTTAATAAGATGTCTCA							
ATGTCAGTTCACCACTTAATAAGCCGTCATCATTTGAAGGTGACCAAGCTGTTTAATAAGATGTCTCA							
X	10	20	30	40	50	60	70
ATGTCAGTTCACCACTTAATAAGCCGTCATCATTTGAAGGTGACCAAGCTGTTTAATAAGATGTCTCA							
ATGTCAGTTCACCACTTAATAAGCCGTCATCATTTGAAGGTGACCAAGCTGTTTAATAAGATGTCTCA							
X	10	20	30	40	50	60	70
ATGTCAGTTCACCACTTAATAAGCCGTCATCATTTGAAGGTGACCAAGCTGTTTAATAAGATGTCTCA							
ATGTCAGTTCACCACTTAATAAGCCGTCATCATTTGAAGGTGACCAAGCTGTTTAATAAGATGTCTCA							
X	10	20	30	40	50	60	70
ATGTCAGTTCACCACTTAATAAGCCGTCATCATTTGAAGGTGACCAAGCTGTTTAATAAGATGTCTCA							
ATGTCAGTTCACCACTTAATAAGCCGTCATCATTTGAAGGTGACCAAGCTGTTTAATAAGATGTCTCA							
X	10	20	30	40	50	60	70
ATGTCAGTTCACCACTTAATAAGCCGTCATCATTTGAAGGTGACCAAGCTGTTTAATAAGATGTCTCA							
ATGTCAGTTCACCACTTAATAAGCCGTCATCATTTGAAGGTGACCAAGCTGTTTAATAAGATGTCTCA							
X	10	20	30	40	50	60	70
ATGTCAGTTCACCACTTAATAAGCCGTCATCATTTGAAGGTGACCAAGCTGTTTAATAAGATGTCTCA							
ATGTCAGTTCACCACTTAATAAGCCGTCATCATTTGAAGGTGACCAAGCTGTTTAATAAGATGTCTCA							
X	10	20	30	40	50	60	70
ATGTCAGTTCACCACTTAATAAGCCGTCATCATTTGAAGGTGACCAAGCTGTTTAATAAGATGTCTCA							
ATGTCAGTTCACCACTTAATAAGCCGTCATCATTTGAAGGTGACCAAGCTGTTTAATAAGATGTCTCA							
X	10	20	30	40	50	60	70
ATGTCAGTTCACCACTTAATAAGCCGTCATCATTTGAAGGTGACCAAGCTGTTTAATAAGATGTCTCA							
ATGTCAGTTCACCACTTAATAAGCCGTCATCATTTGAAGGTGACCAAGCTGTTTAATAAGATGTCTCA							
X	10	20	30	40	50	60	70
ATGTCAGTTCACCACTTAATAAGCCGTCATCATTTGAAGGTGACCAAGCTGTTTAATAAGATGTCTCA							
ATGTCAGTTCACCACTTAATAAGCCGTCATCATTTGAAGGTGACCAAGCTGTTTAATAAGATGTCTCA							
X	10	20	30	40	50	60	70
ATGTCAGTTCACCACTTAATAAGCCGTCATCATTTGAAGGTGACCAAGCTGTTTAATAAGATGTCTCA							
ATGTCAGTTCACCACTTAATAAGCCGTCATCATTTGAAGGTGACCAAGCTGTTTAATAAGATGTCTCA							
X	10	20	30	40	50	60	70
ATGTCAGTTCACCACTTAATAAGCCGTCATCATTTGAAGGTGACCAAGCTGTTTAATAAGATGTCTCA							
ATGTCAGTTCACCACTTAATAAGCCGTCATCATTTGAAGGTGACCAAGCTGTTTAATAAGATGTCTCA							
X	10	20	30	40	50	60	70
ATGTCAGTTCACCACTTAATAAGCCGTCATCATTTGAAGGTGACCAAGCTGTTTAATAAGATGTCTCA							
ATGTCAGTTCACCACTTAATAAGCCGTCATCATTTGAAGGTGACCAAGCTGTTTAATAAGATGTCTCA							
X	10	20	30	40	50	60	70
ATGTCAGTTCACCACTTAATAAGCCGTCATCATTTGAAGGTGACCAAGCTGTTTAATAAGATGTCTCA							
ATGTCAGTTCACCACTTAATAAGCCGTCATCATTTGAAGGTGACCAAGCTGTTTAATAAGATGTCTCA							
X	10	20	30	40	50	60	70
ATGTCAGTTCACCACTTAATAAGCCGTCATCATTTGAAGGTGACCAAGCTGTTTAATAAGATGTCTCA							
ATGTCAGTTCACCACTTAATAAGCCGTCATCATTTGAAGGTGACCAAGCTGTTTAATAAGATGTCTCA							
X	10	20	30	40	50	60	70
ATGTCAGTTCACCACTTAATAAGCCGTCATCATTTGAAGGTGACCAAGCTGTTTAATAAGATGTCTCA							
ATGTCAGTTCACCACTTAATAAGCCGTCATCATTTGAAGGTGACCAAGCTGTTTAATAAGATGTCTCA							
X	10	20	30	40	50	60	70
ATGTCAGTTCACCACTTAATAAGCCGTCATCATTTGAAGGTGACCAAGCTGTTTAATAAGATGTCTCA							
ATGTCAGTTCACCACTTAATAAGCCGTCATCATTTGAAGGTGACCAAGCTGTTTAATAAGATGTCTCA							
X	10	20	30	40	50	60	70
ATGTCAGTTCACCACTTAATAAGCCGTCATCATTTGAAGGTGACCAAGCTGTTTAATAAGATGTCTCA							
ATGTCAGTTCACCACTTAATAAGCCGTCATCATTTGAAGGTGACCAAGCTGTTTAATAAGATGTCTCA							
X	10	20	30	40	50	60	70
ATGTCAGTTCACCACTTAATAAGCCGTCATCATTTGAAGGTGACCAAGCTGTTTAATAAGATGTCTCA							
ATGTCAGTTCACCACTTAATAAGCCGTCATCATTTGAAGGTGACCAAGCTGTTTAATAAGATGTCTCA							
X	10	20	30	40	50	60	70
ATGTCAGTTCACCACTTAATAAGCCGTCATCATTTGAAGGTGACCAAGCTGTTTAATAAGATGTCTCA							
ATGTCAGTTCACCACTTAATAAGCCGTCATCATTTGAAGGTGACCAAGCTGTTTAATAAGATGTCTCA							
X	10	20	30	40	50	60	70
ATGTCAGTTCACCACTTAATAAGCCGTCATCATTTGAAGGTGACCAAGCTGTTTAATAAGATGTCTCA							
ATGTCAGTTCACCACTTAATAAGCCGTCATCATTTGAAGGTGACCAAGCTGTTTAATAAGATGTCTCA							
X	10	20	30	40	50	60	70
ATGTCAGTTCACCACTTAATAAGCCGTCATCATTTGAAGGTGACCAAGCTGTTTAATAAGATGTCTCA							
ATGTCAGTTCACCACTTAATAAGCCGTCATCATTTGAAGGTGACCAAGCTGTTTAATAAGATGTCTCA							
X	10	20	30	40	50	60	70
ATGTCAGTTCACCACTTAATAAGCCGTCATCATTTGAAGGTGACCAAGCTGTTTAATAAGATGTCTCA							
ATGTCAGTTCACCACTTAATAAGCCGTCATCATTTGAAGGTGACCAAGCTGTTTAATAAGATGTCTCA							
X	10	20	30	40	50	60	70
ATGTCAGTTCACCACTTAATAAGCCGTCATCATTTGAAGGTGACCAAGCTGTTTAATAAGATGTCTCA							
ATGTCAGTTCACCACTTAATAAGCCGTCATCATTTGAAGGTGACCAAGCTGTTTAATAAGATGTCTCA							
X	10	20	30	40	50	60	70
ATGTCAGTTCACCACTTAATAAGCCGTCATCATTTGAAGGTGACCAAGCTGTTTAATAAGATGTCTCA							
ATGTCAGTTCACCACTTAATAAGCCGTCATCATTTGAAGGTGACCAAGCTGTTTAATAAGATGTCTCA							
X	10	20	30	40	50	60	70
ATGTCAGTTCACCACTTAATAAGCCGTCATCATTTGAAGGTGACCAAGCTGTTTAATAAGATGTCTCA							
ATGTCAGTTCACCACTTAATAAGCCGTCATCATTTGAAGGTGACCAAGCTGTTTAATAAGATGTCTCA							
X	10	20	30	40	50	60	70
ATGTCAGTTCACCACTTAATAAGCCGTCATCATTTGAAGGTGACCAAGCTGTTTAATAAGATGTCTCA							
ATGTCAGTTCACCACTTAATAAGCCGTCATCATTTGAAGGTGACCAAGCTGTTTAATAAGATGTCTCA							
X	10	20	30	40	50	60	70
ATGTCAGTTCACCACTTAATAAGCCGTCATCATTTGAAGGTGACCAAGCTGTTTAATAAGATGTCTCA							
ATGTCAGTTCACCACTTAATAAGCCGTCATCATTTGAAGGTGACCAAGCTGTTTAATAAGATGTCTCA							
X	10	20	30	40	50	60	70
ATGTCAGTTCACCACTTAATAAGCCGTCATCATTTGAAGGTGACCAAGCTGTTTAATAAGATGTCTCA							
ATGTCAGTTCACCACTTAATAAGCCGTCATCATTTGAAGGTGACCAAGCTGTTTAATAAGATGTCTCA							
X	10	20	30	40	50	60	70
ATGTCAGTTCACCACTTAATAAGCCGTCATCATTTGAAGGTGACCAAGCTGTTTAATAAGATGTCTCA							
ATGTCAGTTCACCACTTAATAAGCCGTCATCATTTGAAGGTGACCAAGCTGTTTAATAAGATGTCTCA							
X	10	20	30	40	50	60	70
ATGTCAGTTCACCACTTAATAAGCCGTCATCATTTGAAGGTGACCAAGCTGTTTAATAAGATGTCTCA							
ATGTCAGTTCACCACTTAATAAGCCGTCATCATTTGAAGGTGACCAAGCTGTTTAATAAGATGTCTCA							
X	10	20	30	40	50	60	70
ATGTCAGTTCACCACTTAATAAGCCGTCATCATTTGAAGGTGACCAAGCTGTTTAATAAGATGTCTCA							
ATGTCAGTTCACCACTTAATAAGCCGTCATCATTTGAAGGTGACCAAGCTGTTTAATAAGATGTCTCA							
X	10	20	30	40	50	60	70
ATGTCAGTTCACCACTTAATAAGCCGTCATCATTTGAAGGTGACCAAGCTGTTTAATAAGATGTCTCA							
ATGTCAGTTCACCACTTAATAAGCCGTCATCATTTGAAGGTGACCAAGCTGTTTAATAAGATGTCTCA							
X	10	20	30	40	50	60	70
ATGTCAGTTCACCACTTAATAAGCCGTCATCATTTGAAGGTGACCAAGCTGTTTAATAAGATGTCTCA							
ATGTCAGTTCACCACTTAATAAGCCGTCATCATTTGAAGGTGACCAAGCTGTTTAATAAGATGTCTCA							
X	10	20	30	40	50	60	70
ATGTCAGTTCACCACTTAATAAGCCGTCATCATTTGAAGGTGACCAAGCTGTTTAATAAGATGTCTCA							
ATGTCAGTTCACCACTTAATAAGCCGTCATCATTTGAAGGTGACCAAGCTGTTTAATAAGATGTCTCA							
X	10	20	30	40	50	60	70
ATGTCAGTTCACCACTTAATAAGCCGTCATCATTTGAAGGTGACCAAGCTGTTTAATAAGATGTCTCA							
ATGTCAGTTCACCACTTAATAAGCCGTCATCATTTGAAGGTGACCAAGCTGTTTAATAAGATGTCTCA							
X	10	20	30	40	50	60	70
ATGTCAGTTCACCACTTAATAAGCCGTCATCATTTGAAGGTGACCAAGCTGTTTAATAAGATGTCTCA							
ATGTCAGTTCACCACTTAATAAGCCGTCATCATTTGAAGGTGACCAAGCTGTTTAATAAGATGTCTCA							
X	10	20	30	40	50	60	70
ATGTCAGTTCACCACTTAATAAGCCGTCATCATTTGAAGGTGACCAAGCTGTTTAATAAGATGTCTCA							
ATGTCAGTTCACCACTTAATAAGCCGTCATCATTTGAAGGTGACCAAGCTGTTTAATAAGATGTCTCA							
X	10	20	30	40	50	60	70
ATGTCAGTTCACCACTTAATAAGCCGTCATCATTTGAAGGTGACCAAGCTGTTTAATAAGATGTCTCA							
ATGTCAGTTCACCACTTAATAAGCCGTCATCATTTGAAGGTGACCAAGCTGTTTAATAAGATGTCTCA							
X	10	20	30	40	50	60	70
ATGTCAGTTCACCACTTAATAAGCCGTCATCATTTGAAGGTGACCAAGCTGTTTAATAAGATGTCTCA							
ATGTCAGTTCACCACTTAATAAGCCGTCATCATTTGAAGGTGACCAAGCTGTTTAATAAGATGTCTCA							

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2003, 09:22:06 ; Search time 43 Seconds

(without alignments)
1146.575 Million cell updates/sec

Title: US-10-081-644-2

Perfect score: 1920
Sequence: 1 MSVPTQCAVIVIEGDKAVVK.....ALTGEGKKNKKNVYVARL 370

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

A_Geneseq_101002:*

1: /SID2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SID2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SID2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SID2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SID2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SID2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SID2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SID2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SID2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SID2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SID2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SID2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SID2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SID2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SID2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SID2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SID2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SID2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SID2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SID2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SID2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SID2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SID2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	310.5	16.2	131	23	ABP10927
2	266.5	13.9	363	21	AAV96748
3	266	13.9	318	21	AAAB27051
4	266	13.9	318	23	AAU80748
5	265	13.8	322	22	AAU35110
6	264	13.8	313	23	ABP47714
7	261	13.6	386	21	AA643467
8	254	13.2	309	21	AA643468
9	254	13.2	386	21	AA622152
10	247	12.9	309	21	AA622153

11	206	10.7	338	22	AAU35351
12	173	9.0	328	23	ABP54030
13	168	8.8	337	22	AA639071
14	168	8.8	337	22	AA679594
15	167	8.7	363	21	AA628581
16	167	8.7	363	21	AA652443
17	167	8.7	366	21	AA628580
18	167	8.7	366	21	AA652442
19	164.5	8.6	324	21	AA639983
20	164.5	8.6	329	21	AA639982
21	164.5	8.6	329	23	ABP29214
22	164.5	8.6	354	21	AA639981
23	162	8.4	262	21	AA605068
24	162	8.4	267	21	AA605067
25	162	8.4	292	21	AA605066
26	158.5	8.3	396	17	AA67569
27	158.5	8.3	413	22	ABP11836
28	155	8.1	117	23	AAU80747
29	155	8.1	334	22	AAU59488
30	155	8.1	334	22	AAU59487
31	153.5	8.0	336	16	ABP38413
32	151.5	7.9	340	23	ABP38413
33	151	7.9	325	21	AA617585
34	151	7.9	325	21	AA650141
35	149.5	7.8	322	20	AA623994
36	149.5	7.8	332	21	AA659181
37	149	7.8	318	22	AA698976
38	148.5	7.7	348	21	AA617994
39	148.5	7.7	348	21	AA627477
40	148.5	7.7	348	23	ABP26966
41	148.5	7.7	351	21	AA627476
42	147.5	7.7	351	21	AA605685
43	147.5	7.7	362	21	AA605684
44	147	7.7	343	21	AA628986
45	145.5	7.6	357	23	ABP28623

ALIGNMENTS

RESULT 1					
ID	ABP10927	standard; Protein; 131 AA.			
XX					
AC	ABP10927;				
DT	25-JUN-2002	(first entry)			
DE					
XX					
Human ORFX protein sequence SEQ ID NO:21836.					
KW	Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;				Enterococcus faeca
KW	hyperproliferative disorder; psoriasis; benign tumour; hemorrhage;				Lactococcus lactis
KW	degenerative disorder; osteoarthritis; neurodegenerative disorder;				C glutamicum prote
KW	cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;				Corynebacterium gl
KW	hypertension; hypothyroidism; cholesterol ester storage disease;				Arabidopsis thalia
KW	immune deficiency; immune disorder; infectious disease;				Arabidopsis thalia
KW	autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;				Arabidopsis thalia
KW	myaschenia gravis.				Arabidopsis thalia
XX					
OS	Homo sapiens.				
XX					
PN	MO200192523-AA2.				
XX					
PD	06-DEC-2001.				
XX					
PF	29-MAY-2001; 2001WO-US10836.				
XX					
PR	30-MAY-2000; 2000US-206132P.				
XX					
PR	29-AUG-2000; 2000US-228716P.				
XX					
PA	(CURA-) CURAGEN CORP.				
XX					
PI	Shinketsu RA, Leach MD;				

XX WPI; 2002-106308/14.
 DR N-PSDB; ABN26679.
 XX
 PT Novel human polypeptides and polynucleotides useful for diagnosing,
 PT preventing and treating cardiovascular disease, neurodegenerative,
 PT hyperproliferative disorders and autoimmune disorders
 XX
 PS Disclosure; SEQ ID 21836; 1037pp; English.
 XX
 CC The present invention describes substantially purified human proteins
 CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1
 CC in the specification). ABN15762 to ABN27252 encode the human ORFX
 CC proteins given in ABP00010 to ABP1500. ORFX proteins are useful for
 CC treating or preventing a pathology associated with an ORFX-associated
 CC disorder in humans, and in the manufacture of a medicament for treating a
 CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide
 CC sequences can be used in gene therapy. ORFX sequences can be used in the
 CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
 CC psoriasis, benign tumors, keloid, degenerative disorders, haemorrhage,
 CC osteoarthritis, neurodegenerative disorders, disorders related to organ
 CC transplantation, cardiovascular diseases, diabetes mellitus, systemic
 CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester
 CC storage disease, various immune deficiencies and disorders, infectious
 CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
 CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
 CC disease and autoimmune inflammatory eye disease. ORFX proteins are also
 CC useful for treating burns, incisions, ulcers, for treating osteoporosis,
 CC bone degenerative disorders, or periodontal disease, and for gut
 CC protection or regeneration and treatment of lung or liver fibrosis,
 CC reperfusion injury in various tissues and conditions resulting from
 CC systemic cytokine damage.
 CC N.B. The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 131 AA;
 XX
 Query Match 16.2%; Score 310.5; DB 23; Length 131;
 Best Local Similarity 49.2%; Pred. No. 2e-20;
 Matches 64; Conservative 16; Mismatches 47; Indels 3; Gaps 2;
 XX
 QY 74 PNA-STDLKVGDFGFGVHGAASQTPKNGAPAEYARVPPLEFK--SNLTHSTADEISEG 130
 DB 2 PNDVAAAFALGDYIVYVIGASVAFPPNGAFPAEYSALISSETAYKPAEFRLCKDKLPEG 61
 QY 131 PVKMFESAALPVSLTAGVSLCHILGSKMEHPSSTQHHPLLIWGGATAVGQOLIOVA 190
 DB 62 PVSLBGAVALPVSLTAGVSLCHILGSKMEHPSSTQHHPLLIWGGATAVGQOLIOVA 121
 QY 191 KHINAYTKIV 200
 DB 122 KKLNGFSKII 131
 XX
 RESULT 2
 AAAY96748
 ID AAAY96748 standard; Protein; 363 AA.
 XX
 AC AAAY96748;
 XX
 DT 09-OCT-2000 (first entry)
 DE
 XX A. terreus LovC dehydrogenase.
 XX
 KM Lovastatin; D4B segment; monacolin J; dehydrogenase; LovC; anti-lipemic;
 KM HMG-CoA reductase inhibitor; anti-hypercholesterolaemic; anti-fungal.
 XX
 OS Aspergillus terreus.
 XX
 FN MO200037629-A2.
 XX
 PT 29-JUN-2000.

XX
 PF 13-DEC-1999; 99WO-US29583.
 XX
 PR 18-DEC-1998; 98US-0215694.
 XX
 PA (WISC) WISCONSIN ALUMNI RES FOUND.
 XX
 PI Hutchinson RC, Kennedy J, Park C;
 XX
 DR WPI; 2000-442660/38.
 DR N-PSDB; AAAS1300.
 XX
 PT Increasing lovastatin or monacolin J production in an organism, for use
 PT as antihypercholesterolemic or antifungal agents, comprises
 PT transforming the organism with a D4B segment
 XX
 PS Disclosure; Page 62-63; 116pp; English.
 XX
 CC The proteins shown in AAAY96744-60 are encoded by 17 genes from a
 CC cluster in Aspergillus terreus (ATCC 20542), which flank the NPKS
 CC (nonaketide polyketide synthase) gene, which is known to be required for
 CC lovastatin production. The NPKS gene is contained within the context of
 CC the entire gene cluster but is not indicated here (see US5744350). The
 CC genes and proteins are named "ORF" or "Lov", where "Lov" signifies genes
 CC shown to be essential for lovastatin production. The portion of the gene
 CC cluster between ORF1 and the mid-region of LovF is referred to as the
 CC "D4B segment". Increasing lovastatin, or monacolin J, production in a
 CC lovastatin-producing organism, comprises transforming the organism with
 CC a D4B segment, and expressing it (claimed). Lovastatin will also be
 CC produced in non-lovastatin producing organisms (e.g. A. nidulans) by
 CC transformation with the D4B segment and the entire LovF gene. The
 CC methods are used to increase biosynthetic production of lovastatin (with
 CC an at least 5-fold increase) which is an anti-hypercholesterolaemic
 CC agent, and also has some anti-fungal activity. Lovastatin inhibits the
 CC conversion of hydroxymethylglutarylcoenzyme A (HMG-CoA) into mevalonate
 CC by HMG-CoA reductase. The methods can also be used to increase production
 CC of monacolin J (claimed), which has anti-fungal activity.
 XX
 SQ Sequence 363 AA;
 XX
 Query Match 13.9%; Score 266.5; DB 21; Length 363;
 Best Local Similarity 26.3%; Pred. No. 9.4e-16;
 Matches 103; Conservative 69; Mismatches 156; Indels 63; Gaps 17;
 XX
 QY 4 PTTOKAVII-ESDKAVKTVDSVPEIKEGTALVYKVAAGNPTDWMHAIKIGPGSILG 62
 DB 9 PQQTALVTNDBDEVTVMAAPCPMLPRDQVVRVEAVAINSDPK-NMGQFATPMAFLG 67
 QY 63 CDLAGTVKLGPRASIDLVKVGDTGFGFVNGASQTPKNGAPAEY--ARVPPLEFKYN 118
 DB 68 TVAGTVAVVAGSDV-THIOGDRVYGAONEMCPRTPDQAFSQYTVTRGVA----- 119
 QY 119 LTHSTADEISEGVPKMFESAALPVSLTAGVSLCHILGSKMEHPSST--POHTHP--L 173
 DB 120 -----KIPKG--LSFEQALALPAGISITAGLAM-KLGLPLP-SPSADQPTHSKPYV 168
 QY 174 LIWGGATAVGQOLIOVAKHINAYTKIVYVASKHEKILKSYGADVDFVHDAGVLEIQKS 233
 DB 169 LVYGGSTATAVATVMQLR-LSGYIPLAT-CPHNFDLAKSGAEVDFRABNLAQITRT 226
 QY 234 -KYPNIQHYIDAVGSEDSIPEAVKTYTADSLPATLLEVVMTLESTPEE-IRDNVKTIDT 291
 DB 227 YTKNNILRYALDCTIVNESTTFCPFAIGRAGHY-----VSLNPPPEHATKMTWTDWT 280
 QY 292 LLYRASGCEILGATRF-----PASPEYHEATVKFVKFNLPHLNGDIIHNNI 339
 DB 281 -----LGPITFGSGTWPAPYGRPGSEFERQFEDMLRIAGOLVEDGRVLVHNL 329
 QY 340 KYFNSGLDVPALTEIGIEKGNKKNVXYARL 370
 DB 330 RVVQGGFHHIKQGMELVRKGLSGELVYRL 360

QY 226 GVIEQIKSKYKNLOHVIDAVGSEDSIPKAYK 256
 Db 212 N-FEDLPEKF---DVYVDYTGIEIRAKAVK 238

RESULT 5
 ID AAU35110 standard; Protein; 322 AA.
 AAU35110;

AC AAU35110;
 DT 13-FEB-2002 (first entry)

DE Enterococcus faecalis cellular proliferation protein #397.

KM Antisense; prokaryotic cellular proliferation protein;
 KW antibiotic; antibacterial; drug design.

XX Enterococcus faecalis.

MO200170955-A2.

XX 27-SEP-2001.

PF 21-MAR-2001; 2001MO-US09180.

XX 21-MAR-2000; 2000US-191078P.

PR 23-MAY-2000; 2000US-206848P.

PR 26-MAY-2000; 2000US-207272P.

PR 23-OCT-2000; 2000US-242578P.

PR 27-NOV-2000; 2000US-253625P.

PR 22-DEC-2000; 2000US-257931P.

PR 16-FEB-2001; 2001US-269308P.

XX (ELIIR-) ELIIR PHARM INC.

XX Haselbeck R, Ohlsen KL, Zyckind JW, Wall D, Trawick JD, Carr GJ,
 PI Yamamoto RT, Xu HH;

DR N-PSDB; AASS2969.

XX WPI; 2001-611495/70.

PT New polynucleotides for the identification and development of
 XX antibiotics, comprise sequences of antisense nucleic acids -

PS Example 3; Seq ID No 10703; 511pp; English.

XX The invention relates to antisense inhibitors of genes essential to
 CC prokaryotic cellular proliferation, their use in identifying the
 CC genes, their use in the discovery of novel antibiotics, the essential
 CC genes themselves and the encoded proteins. The prokaryotes used are
 CC *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella*
 CC *pneumoniae*, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The
 CC invention is also useful for the identification of potential new targets
 CC for antibiotic development. The antisense nucleic acids can also be used
 CC to identify proteins used in proliferation, to express these proteins,
 CC and to obtain antibodies capable of binding to the expressed proteins.
 CC The proteins can be used to screen compounds in rational drug discovery
 CC programmes. The antisense nucleic acid sequence is also useful to screen
 CC for homologous nucleic acids which are required for cell proliferation in
 CC a wide variety of organisms. The present sequence represents an
 CC essential prokaryotic cellular proliferation protein.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 322 AA;

Query Match 13.8%; Score 265; DB 22; Length 322;
 Best Local Similarity 30.4%; Pred. No. 1, 1e-15;
 Matches 106; Conservative 42; Mismatches 125; Indels 76; Gaps 15;

QY 8 KAVIIE--GDKAUV-KTDVSVBELKGNALVKEVAAGNPDTMKHIAVYKIGEG----- 58
 Db 11 KAVINQVSKSEVLEAEAVTLPELSEHGVLVKEVATSIPIWMKL-----REGYLKQMF 64

QY 59 -----SILGCDIAGTVVKKLGPNASTDLYKVDGTFGFGVHGAASQTPDPNGAFAFYARYPTL 113

Db 65 DMSFPIILGMDVAGVTEVGSQV-TDWDQGDVVF-----ARPETRFGTYAEV--T 112

QY 114 FYKSNLTHSTADEISEGPVKFESASALPVSILITAGVSLCHHLGSKMEWHPESTPOHTPL 173

Db 113 IYDDHILAKIPETIS-----FEAAAVPLAGLTMOALFDGHKL-----EGEV 157

QY 174 LIWGATVAGQQLIQVAKHINAYTKIVVASKKHEKLKSYGADVDYDHAGVIEQI-- 231

Db 158 LIHAGAGVGTYAIIQAKKAGAH--VITPASAKNHALLKKIGADEVIDHTTNFAEVLTD 215

QY 232 -----KSKY---PNIQHVDAVGSSEDSIPKAYK-VTADSLPATLLEVPMT 273

Db 216 VDLVFDPTMGGEVQKNSFDVIAKPNTRGLVSIIVEIKQLAEKNVSAESI--WLPNGEO 272

QY 274 IESIPKIRKDNVKIDITLLYRASGOEILGATRPASPEVHEATVKEFV 322

Db 273 LQKIDLMAGKVKATIGEVPPFSRQ---GIYDAHALSETHHAGKIV 317

RESULT 6

ABBA47714
 ID ABBA47714 standard; Protein; 313 AA.

XX ABBA47714;

AC 05-FEB-2002 (first entry)

XX Listeria monocytogenes protein #418.

XX Anti-bacterial; gene therapy; vaccine; biosynthesis; biodegradation;
 KW vitamin B12; Bacterial infection; disease.

OS Listeria monocytogenes.

XX MO200177335-A2.

XX 18-OCT-2001.

XX 11-APR-2001; 2001MO-FR01118.

XX 11-APR-2000; 2000FR-0004629.

XX (INSP) INST PASTEUR.

XX Buchrieser C, Frangaul L, Couve E, Ruenick C, Fathi H, Deboux P,
 PI Dussurget O, Cherouani F, Nedjari H, Glaser P, Kunst F, Cossart P;
 PI Daniels J, Goebel W, Kreft J, Kuhn W, Ng E, Vazquez-Boland JA;
 PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A,
 PI Chakraborty T, Doman E, Hain T, Berche P, Chardot A, Durant L;
 PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;
 PI Madueno E, De Pablos B, Wehland J, Kaerst U, Entian K, Haut J;
 PI Rose M, Voss H;
 XX WPI; 2002-010914/01.

XX Genomic sequence for *Listeria monocytogenes*, useful e.g. for treatment
 PT and prevention of *Listeria* and related bacterial infections, and
 PT related polypeptides -

PS Claim 6; SEQ ID No 419; 192pp; French.

XX The present invention relates to the genome sequence of *Listeria*
 CC monocytogenes EGD-e (see ABBA03041). The genome sequence and fragments of
 CC it are useful for selecting probes and primers for detecting genes in *L.*
 CC monocytogenes and related organisms, and for studying genetic
 CC polymorphisms and other genomes. The present sequence is a protein

PR	19-UUTL-1999	99US-0144331
PR	19-UUTL-1999	99US-0144332
PR	19-UUTL-1999	99US-0144333
PR	19-UUTL-1999	99US-0144334
PR	19-UUTL-1999	99US-0144335
PR	20-UUTL-1999	99US-0144352
PR	20-UUTL-1999	99US-0144353
PR	20-UUTL-1999	99US-0144684
PR	21-UUTL-1999	99US-0144814
PR	21-UUTL-1999	99US-0145088
PR	21-UUTL-1999	99US-0145089
PR	22-UUTL-1999	99US-0145085
PR	22-UUTL-1999	99US-0145087
PR	22-UUTL-1999	99US-0145089
PR	23-UUTL-1999	99US-0145152
PR	23-UUTL-1999	99US-0145218
PR	23-UUTL-1999	99US-0145524
PR	26-UUTL-1999	99US-0145576
PR	27-UUTL-1999	99US-0145513
PR	27-UUTL-1999	99US-0145918
PR	27-UUTL-1999	99US-0145919
PR	28-UUTL-1999	99US-0145951
PR	02-AUG-1999	99US-0146386
PR	02-AUG-1999	99US-0146388
PR	02-AUG-1999	99US-0146389
PR	03-AUG-1999	99US-0147038
PR	04-AUG-1999	99US-0147034
PR	04-AUG-1999	99US-0147192
PR	05-AUG-1999	99US-0147193
PR	05-AUG-1999	99US-0147760
PR	06-AUG-1999	99US-0147763
PR	06-AUG-1999	99US-0147810
PR	09-AUG-1999	99US-0147933
PR	09-AUG-1999	99US-0147935
PR	10-AUG-1999	99US-0148171
PR	11-AUG-1999	99US-0148319
PR	12-AUG-1999	99US-0148341
PR	13-AUG-1999	99US-0148684
PR	13-AUG-1999	99US-0148685
PR	16-AUG-1999	99US-0149168
PR	17-AUG-1999	99US-0149375
PR	18-AUG-1999	99US-0149426
PR	20-AUG-1999	99US-0149722
PR	20-AUG-1999	99US-0149723
PR	20-AUG-1999	99US-0149902
PR	22-AUG-1999	99US-0149929
PR	23-AUG-1999	99US-0149930
PR	25-AUG-1999	99US-0150566
PR	26-AUG-1999	99US-0150884
PR	27-AUG-1999	99US-0151065
PR	27-AUG-1999	99US-0151066
PR	27-AUG-1999	99US-0151080
PR	30-AUG-1999	99US-0151013
PR	31-AUG-1999	99US-0151438
PR	01-SEP-1999	99US-0151530
PR	07-SEP-1999	99US-0152373
PR	10-SEP-1999	99US-0153070
PR	13-SEP-1999	99US-0153758
PR	15-SEP-1999	99US-0154018
PR	16-SEP-1999	99US-0154019
PR	20-SEP-1999	99US-0154779
PR	22-SEP-1999	99US-0155139
PR	23-SEP-1999	99US-0155846
PR	24-SEP-1999	99US-0155659
PR	26-SEP-1999	99US-0156458
PR	29-SEP-1999	99US-0156596
PR	04-OCT-1999	99US-0157117
PR	05-OCT-1999	99US-0157753
PR	06-OCT-1999	99US-0157865
PR	07-OCT-1999	99US-0158029
PR	08-OCT-1999	99US-0158323
PR	12-OCT-1999	99US-0158369

PR	13-OCT-1999	99US-01592924
PR	13-OCT-1999	99US-01592934
PR	13-OCT-1999	99US-01592855
PR	14-OCT-1999	99US-01593229
PR	14-OCT-1999	99US-01593930
PR	14-OCT-1999	99US-01593331
PR	14-OCT-1999	99US-01596637
PR	14-OCT-1999	99US-01596588
PR	18-OCT-1999	99US-01595854
PR	21-OCT-1999	99US-01607441
PR	21-OCT-1999	99US-01607768
PR	21-OCT-1999	99US-01607657
PR	21-OCT-1999	99US-01607770
PR	21-OCT-1999	99US-01607700
PR	21-OCT-1999	99US-01608154
PR	21-OCT-1999	99US-01608815
PR	22-OCT-1999	99US-01609881
PR	22-OCT-1999	99US-01609881
PR	22-OCT-1999	99US-01609889
PR	22-OCT-1999	99US-01614005
PR	22-OCT-1999	99US-01614005
PR	22-OCT-1999	99US-01614005
PR	22-OCT-1999	99US-01614005
PR	22-OCT-1999	99US-01613569
PR	22-OCT-1999	99US-01613660
PR	22-OCT-1999	99US-01613661
PR	28-OCT-1999	99US-01619270
PR	28-OCT-1999	99US-01619393
PR	28-OCT-1999	99US-01621442

Query Match	13.6%;	Score 261;	DB 21;	Length 386;
Best Local Similarity	-32.2%;	Pred. No. 3.3e-15;		
Matches	88;	Conservative	37;	Mismatches 104;
				Indels 44;
				Gaps 11

Qy		SVPTTKQAIIIG-----DKAVYKTDVSPBELKEGTALVKVEAAGNPTMKHIAVYKI---	54
Db		73 SLEPKENKAANWVS DYSGGVVLKLESNTIVPELKEQOVLIKVAALAALNVDMRRQGFKAT	132
Qy		55 -GPEGSIILGCDIAGTVVYLKGEPNASITDLKVGDGTGFGFPHGASQTDPKN-GAFPAEAYRYP	112
Db		133 DSELPFVPBGIDVAGVVAKVG-SAYVDLKEGBEVYANISERALEBPGFSLAEYAVEEK	191
Qy		113 L--FYKSNIUHTSHADEISEBPVKPFESNASLPSVSLTTAGVSLCHHLGSKMEHPSTPQHT	170
Db		192 LALKRKNI-----DPAQAAGPLAIETADGGLV-----REFSAG-----	227
Qy		171 HPIILWGATTAAGQOLIQAANKHINAAYKIYITVASKEHEKLKSYGADVDFDYHDAGYTIO	230
Db		228 KSLIVNGAGGGSIVTOLAKHVYGASKAATASTEKLEIVRSIGLDLAIDYTKEN-IED	286
Qy		231 IKSRYPNLOHVIDAVGSEDS----IPRAKYKTA	259
Db		287 LPDKY---DVFPDAIGMCDKAAVKVIKEGGKVA	316

RESULT 8
AAG43468
ID AAG43468 standard; Protein; 309 AA.
XX AC AAG43468;
XX DT 18-OCT-2000 (first entry)
XX EE Arabidopsis thaliana protein fragment SEQ ID NO: 54335.
XX KW Protein identification; signal transduction pathway; metabolic pathway;
XX KM hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX OS Arabidopsis thaliana.
XX PN EPI033405-AZ.
XD 06-SEP-2000.

XX - 25-FEB-2000; 2000EP-0301439.
PF XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 07-MAY-1999; 99US-0132486.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.

PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144332.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145919.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147033.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149366.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153759.
PR 13-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155113.
PR 23-SEP-1999; 99US-0155486.

PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 13.2%; Score 254; DB 21; Length 309;
Best Local Similarity 32.8%; Pred. No. 1e-14;
Matches 84; Conservative 35; Mismatches 97; Indels 40; Gaps 10;

QY 15 DKAVKTDVSPPEIKETALVKYEAAGNPTDMKHIYK-----GPGSILGDIAGTVV 70
DB 13 DVKTESNIVVPEIKEDQVLIKVAALNPDVDAKROGKFRKARDSPFVPGYDVAGVVV 72
QY 71 KLGPNASTDVKVGTGFGFVHAGASQTDPKK-GAFAEYARYVPL--FYKSNLTHSTADEI 127
DB 73 KVG-SAVKDLKEDDEVYANVSEKALEGPKQFGSLAETVAVEKLLAKPKNI----- 123
QY 128 SEGPRVKNFESASLPSVLTAGVSLCHHLGSKMWEHPSTPOHTHPLLIMGATAVGQOLI 187
DB 124 -----DFAQAGLPALIEFTADEGLV-----RTEFSAG-----KSLVLNAGAGVGLVI 167
QY 188 QVAKHNATYKITYVASKKREKLLKSGADVDVPHNAGVTEQIKSKYPLQHVIVAVGS 247
DB 168 QLKAKHYAGSKVATSTKTELEVRSLGADLAIDYTKEN-IEDLPDKY---DVVFAIAGM 223
QY 248 EDS-----IPEAYKYTA 259
DB 224 CDKAKVYIKEGGKVVV 239

RESULT 9
AAG22152
ID AAG22152 standard; Protein; 386 AA.

XX AAG22152;
XX 17-OCT-2000 (first entry)
XX Arabidopsis thaliana protein fragment SEQ ID NO: 24972.
DB
XX

KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX Arabidopsis thaliana.
XX EPI033405-A2.
XX 06-SEP-2000.
XX 25-FEB-2000; 2000EP-0301439.
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 25-MAR-1999; 99US-0126264.
XX 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127462.
XX 06-APR-1999; 99US-0128234.
XX 08-APR-1999; 99US-0128714.
XX 16-APR-1999; 99US-0129845.
XX 19-APR-1999; 99US-0130077.
XX 21-APR-1999; 99US-0130449.
XX 23-APR-1999; 99US-0130510.
XX 23-APR-1999; 99US-0130891.
XX 28-APR-1999; 99US-0131449.
XX 30-APR-1999; 99US-0132048.
XX 30-APR-1999; 99US-0132407.
XX 04-MAY-1999; 99US-0132484.
XX 05-MAY-1999; 99US-0132485.
XX 06-MAY-1999; 99US-0132486.
XX 06-MAY-1999; 99US-0132487.
XX 07-MAY-1999; 99US-0132863.
XX 11-MAY-1999; 99US-0134256.
XX 14-MAY-1999; 99US-0134218.
XX 14-MAY-1999; 99US-0134221.
XX 14-MAY-1999; 99US-0134370.
XX 18-MAY-1999; 99US-0134768.
XX 19-MAY-1999; 99US-0134941.
XX 20-MAY-1999; 99US-0135124.
XX 21-MAY-1999; 99US-0135353.
XX 24-MAY-1999; 99US-0135629.
XX 25-MAY-1999; 99US-0136021.
XX 27-MAY-1999; 99US-0136392.
XX 28-MAY-1999; 99US-0136782.
XX 01-JUN-1999; 99US-0137222.
XX 03-JUN-1999; 99US-0137528.
XX 04-JUN-1999; 99US-0137502.
XX 07-JUN-1999; 99US-0137724.
XX 08-JUN-1999; 99US-0138094.
XX 10-JUN-1999; 99US-0138540.
XX 10-JUN-1999; 99US-0138847.
XX 14-JUN-1999; 99US-0139119.
XX 16-JUN-1999; 99US-0139452.
XX 16-JUN-1999; 99US-0139453.
XX 17-JUN-1999; 99US-0139454.
XX 18-JUN-1999; 99US-0139454.
XX 18-JUN-1999; 99US-0139455.
XX 18-JUN-1999; 99US-0139456.
XX 18-JUN-1999; 99US-0139457.
XX 18-JUN-1999; 99US-0139458.
XX 18-JUN-1999; 99US-0139459.
XX 18-JUN-1999; 99US-0139460.
XX 18-JUN-1999; 99US-0139461.
XX 18-JUN-1999; 99US-0139462.
XX 18-JUN-1999; 99US-0139463.
XX 18-JUN-1999; 99US-0139464.
XX 18-JUN-1999; 99US-0139465.
XX 21-JUN-1999; 99US-0139817.
XX 22-JUN-1999; 99US-0139899.
XX 23-JUN-1999; 99US-0140353.

ID	Accession	Standard	Protein	Size (AA)
ADG22153	standard	Protein	309	AA
XX	AA022153			
AC	AA022153			
XX				
DT	17-OCT-2000	(first entry)		
XX				
DE	Arabidopsis thaliana protein fragment	SEQ ID NO: 24973		
XX				
KW	Protein identification; signal transduction pathway; metabolic pathway; hybridisation assay; genetic mapping; gene expression control; promoter; termination sequence.			
KM				
XX				
OS	Arabidopsis thaliana.			
XX				
PN	EP1033405-A2.			
XX				
PD	06-SEP-2000.			
XX				
25-FEB-2000;	200SEP-0301439.			
25-FEB-1999;	99US-0121825.			
PR	05-MAR-1999;	99US-0123180.		
PR	09-MAR-1999;	99US-0123548.		
PR	23-MAR-1999;	99US-0125788.		
PR	25-MAR-1999;	99US-0126264.		
PR	29-MAR-1999;	99US-0126785.		
PR	01-APR-1999;	99US-0127462.		
PR	06-APR-1999;	99US-0128234.		
PR	06-APR-1999;	99US-0128714.		
PR	16-APR-1999;	99US-0129845.		
PR	19-APR-1999;	99US-0130077.		
PR	21-APR-1999;	99US-0130449.		
PR	23-APR-1999;	99US-0130510.		
PR	23-APR-1999;	99US-0130891.		
PR	28-APR-1999;	99US-0131449.		
PR	30-APR-1999;	99US-0132048.		
PR	30-APR-1999;	99US-0132407.		
PR	05-MAY-1999;	99US-0132484.		
PR	05-MAY-1999;	99US-0132485.		
PR	06-MAY-1999;	99US-0132486.		
PR	06-MAY-1999;	99US-0132487.		
PR	07-MAY-1999;	99US-0132863.		
PR	11-MAY-1999;	99US-0134256.		
PR	14-MAY-1999;	99US-0134218.		
PR	14-MAY-1999;	99US-0134219.		
PR	14-MAY-1999;	99US-0134221.		
PR	14-MAY-1999;	99US-0134370.		
PR	18-MAY-1999;	99US-0134768.		
PR	19-MAY-1999;	99US-0134941.		
PR	20-MAY-1999;	99US-0135124.		
PR	21-MAY-1999;	99US-0135153.		
PR	24-MAY-1999;	99US-0135629.		
PR	25-MAY-1999;	99US-0136021.		
PR	27-MAY-1999;	99US-0136392.		
PR	28-MAY-1999;	99US-0136382.		
PR	01-JUN-1999;	99US-0137222.		
PR	03-JUN-1999;	99US-0137528.		
PR	04-JUN-1999;	99US-0137502.		
PR	07-JUN-1999;	99US-0137724.		
PR	08-JUN-1999;	99US-0138094.		
PR	10-JUN-1999;	99US-0138540.		
PR	10-JUN-1999;	99US-0138647.		
PR	14-JUN-1999;	99US-0139119.		
PR	16-JUN-1999;	99US-0139452.		
PR	16-JUN-1999;	99US-0139453.		
PR	17-JUN-1999;	99US-0139492.		
PR	17-JUN-1999;	99US-0139454.		
PR	18-JUN-1999;	99US-0139455.		
PR	18-JUN-1999;	99US-0139456.		
PR	18-JUN-1999;	99US-0139457.		
PR	18-JUN-1999;	99US-0139458.		
PR	18-JUN-1999;	99US-0139459.		
PR	18-JUN-1999;	99US-0139459.		
PR	18-JUN-1999;	99US-0139459.		
PR	18-JUN-1999;	99US-0139460.		
PR	18-JUN-1999;	99US-0139461.		
PR	18-JUN-1999;	99US-0139462.		
PR	18-JUN-1999;	99US-0139463.		
PR	18-JUN-1999;			

PR	23-AUG-1999;	99US-0149902.
PR	23-AUG-1999;	99US-0149930.
PR	25-AUG-1999;	99US-0150566.
PR	26-AUG-1999;	99US-0150884.
PR	27-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-0151066.
PR	27-AUG-1999;	99US-0151080.
PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151338.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0158293.
PR	13-OCT-1999;	99US-0159294.
PR	13-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.

[illegible]

Db 168 QAKHVVYGAASKAATSTGCKELVRLSLGDLAIDYTKEN-IEDLPDKY---DVVFDAIGM 223

OY 248 EDS---IPEAYKYTA 259

Db 224 CDKAVKVIIEGGKRYVA 239

RESULT 11

ID	AAU35351 standard; Protein; 338 AA.
1	MA
2	MA
3	MA
4	MA
5	MA
6	MA
7	MA
8	MA
9	MA
10	MA
11	MA
12	MA
13	MA
14	MA
15	MA
16	MA
17	MA
18	MA
19	MA
20	MA
21	MA
22	MA
23	MA
24	MA
25	MA
26	MA
27	MA
28	MA
29	MA
30	MA
31	MA
32	MA
33	MA
34	MA
35	MA
36	MA
37	MA
38	MA
39	MA
40	MA
41	MA
42	MA
43	MA
44	MA
45	MA
46	MA
47	MA
48	MA
49	MA
50	MA
51	MA
52	MA
53	MA
54	MA
55	MA
56	MA
57	MA
58	MA
59	MA
60	MA
61	MA
62	MA
63	MA
64	MA
65	MA
66	MA
67	MA
68	MA
69	MA
70	MA
71	MA
72	MA
73	MA
74	MA
75	MA
76	MA
77	MA
78	MA
79	MA
80	MA
81	MA
82	MA
83	MA
84	MA
85	MA
86	MA
87	MA
88	MA
89	MA
90	MA
91	MA
92	MA
93	MA
94	MA
95	MA
96	MA
97	MA
98	MA
99	MA
100	MA
101	MA
102	MA
103	MA
104	MA
105	MA
106	MA
107	MA
108	MA
109	MA
110	MA
111	MA
112	MA
113	MA
114	MA
115	MA
116	MA
117	MA
118	MA
119	MA
120	MA
121	MA
122	MA
123	MA
124	MA
125	MA
126	MA
127	MA
128	MA
129	MA
130	MA
131	MA
132	MA
133	MA
134	MA
135	MA
136	MA
137	MA
138	MA
139	MA
140	MA
141	MA
142	MA
143	MA
144	MA
145	MA
146	MA
147	MA
148	MA
149	MA
150	MA
151	MA
152	MA
153	MA
154	MA
155	MA
156	MA
157	MA
158	MA
159	MA
160	MA
161	MA
162	MA
163	MA
164	MA
165	MA
166	MA
167	MA
168	MA
169	MA
170	MA
171	MA
172	MA
173	MA
174	MA
175	MA
176	MA
177	MA
178	MA
179	MA
180	MA
181	

AC AAU35351;

DT 14-FEB-2002 (first entry)

DE Enterococcus faecalis cellular proliferation protein #638.

KW Antisense; prokaryotic cellular proliferation protein;

2000

[illegible]

2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525
 526

2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020 2021 2022 2023 2024 2025 2026 2027 2028 2029 2030 2031 2032 2033 2034 2035 2036 2037 2038 2039 2040 2041 2042 2043 2044 2045 2046 2047 2048 2049 2050 2051 2052 2053 2054 2055 2056 2057 2058 2059 2060 2061 2062 2063 2064 2065 2066 2067 2068 2069 2070 2071 2072 2073 2074 2075 2076 2077 2078 2079 2080 2081 2082 2083 2084 2085 2086 2087 2088 2089 2090 2091 2092 2093 2094 2095 2096 2097 2098 2099 2100 2101 2102 2103 2104 2105 2106 2107 2108 2109 2110 2111 2112 2113 2114 2115 2116 2117 2118 2119 2120 2121 2122 2123 2124 2125 2126 2127 2128 2129 2130 2131 2132 2133 2134 2135 2136 2137 2138 2139 2140 2141 2142 2143 2144 2145 2146 2147 2148 2149 2150 2151 2152 2153 2154 2155 2156 2157 2158 2159 2160 2161 2162 2163 2164 2165 2166 2167 2168 2169 2170 2171 2172 2173 2174 2175 2176 2177 2178 2179 2180 2181 2182 2183 2184 2185 2186 2187 2188 2189 2190 2191 2192 2193 2194 2195 2196 2197 2198 2199 2200 2201 2202 2203 2204 2205 2206 2207 2208 2209 2210 2211 2212 2213 2214 2215 2216 2217 2218 2219 2220 2221 2222 2223 2224 2225 2226 2227 2228 2229 2230 2231 2232 2233 2234 2235 2236 2237 2238 2239 2240 2241 2242 2243 2244 2245 2246 2247 2248 2249 2250 2251 2252 2253 2254 2255 2256 2257 2258 2259 2260 2261 2262 2263 2264 2265 2266 2267 2268 2269 2270 2271 2272 2273 2274 2275 2276 2277 2278 2279 2280 2281 2282 2283 2284 2285 2286 2287 2288 2289 2290 2291 2292 2293 2294 2295 2296 2297 2298 2299 2300 2301 2302 2303 2304 2305 2306 2307 2308 2309 2310 2311 2312 2313 2314 2315 2316 2317 2318 2319 2320 2321 2322 2323 2324 2325 2326 2327 2328 2329 2330 2331 2332 2333 2334 2335 2336 2337 2338 2339 2340 2341 2342 2343 2344 2345 2346 2347 2348 2349 2350 2351 2352 2353 2354 2355 2356 2357 2358 2359 2360 2361 2362 2363 2364 2365 2366 2367 2368 2369 2370 2371 2372 2373 2374 2375 2376 2377 2378 2379 2380 2381 2382 2383 2384 2385 2386 2387 2388 2389 2390 2391 2392 2393 2394 2395 2396 2397 2398 2399 2400 2401 2402 2403 2404 2405 2406 2407 2408 2409 2410 2411 2412 2413 2414 2415 2416 2417 2418 2419 2420 2421 2422 2423 2424 2425 2426 2427 2428 2429 2430 2431 2432 2433 2434 2435 2436 2437 2438 2439 2440 2441 2442 2443 2444 2445 2446 2447 2448 2449 2450 2451 2452 2453 2454 2455 2456 2457 2458 2459 2460 2461 2462 2463 2464 2465 2466 2467 2468 2469 2470 2471 2472 2473 2474 2475 2476 2477 2478 2479 2480 2481 2482 2483 2484 2485 2486 2487 2488 2489 2490 2491 2492 2493 2494 2495 2496 2497 2498 2499 2500 2501 2502 2503 2504 2505 2506 2507 2508 2509 2510 2511 2512 2513 2514 2515 2516 2517 2518 2519 2520 2521 2522 2523 2524 2525 2526 2527 2528 2529 2530 2531 2532 2533 2534 2535 2536 2537 2538 2539 2540 2541 2542 2543 2544 2545 2546 2547 2548 2549 2550 2551 2552 2553 2554 2555 2556 2557 2558 2559 2560 2561 2562 2563 2564 2565 2566 2567 2568 2569 2570 2571 2572 2573 2574 2575 2576 2577 2578 2579 2580 2581 2582 2583 2584 2585 2586 2587 2588 2589 2590 2591 2592 2593 2594 2595 2596 2597 2598 2599 2600 2601 2602 2603 2604 2605 2606 2607 2608 2609 2610 2611 2612 2613 2614 2615 2616 2617 2618 2619 2620 2621 2622 2623 2624 2625 2626 2627 2628 2629 2630 2631 2632 2633 2634 2635 2636 2637 2638 2639 2640 2641 2642 2643 2644 2645 2646 2647 2648 2649 2650 2651 2652 2653 2654 2655 2656 2657 2658 2659 2660 2661 2662 2663 2664 2665 2666 2667 2668 2669 2670 2671 2672 2673 2674 2675 2676 2677 2678 2679 2680 2681 2682 2683 2684 2685 2686 2687 2688 2689 2690 2691 2692 2693 2694 2695 2696 2697 2698 2699 2700 2701 2702 2703 2704 2705 2706 2707 2708 2709 2710 2711 2712 2713 2714 2715 2716 2717 2718 2719 2720 2721 2722 2723 2724 2725 2726 2727 2728 2729 2730 2731 2732 2733 2734 2735 2736 2737 2738 2739 2740 2741 2742 2743 2744 2745 2746 2747 2748 2749 2750 2751 2752 2753 2754 2755 2756 2757 2758 2759 2760 2761 2762 2763 2764 2765 2766 2767 2768 2769 2770 2771 2772 2773 2774 2775 2776 2777 2778 2779 2780 2781 2782 2783 2784 2785 2786 2787 2788 2789 2790 2791 2792 2793 2794 2795 2796 2797 2798 2799 2800 2801 2802 2803 2804 2805 2806 2807 2808 2809 2810 2811 2812 2813 2814 2815 2816 2817 2818

2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525
 526

PR 23-MAY-2000; 2000US-206848P

PR 23-OCT-2000; 2000US-242578P

PR 22-DEC-2000; 2000US-257931P

XX
 (P.T. 1) P.T. 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 223, 224, 225, 226, 227, 228, 229, 230, 231, 232, 233, 234, 235, 236, 237, 238, 239, 240, 241, 242, 243, 244, 245, 246, 247, 248, 249, 250, 251, 252, 253, 254, 255, 256, 257, 258, 259, 260, 261, 262, 263, 264, 265, 266, 267, 268, 269, 270, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, 285, 286, 287, 288, 289, 290, 291, 292, 293, 294, 295, 296, 297, 298, 299, 300, 301, 302, 303, 304, 305, 306, 307, 308, 309, 310, 311, 312, 313, 314, 315, 316, 317, 318, 319, 320, 321, 322, 323, 324, 325, 326, 327, 328, 329, 330, 331, 332, 333, 334, 335, 336, 337, 338, 339, 340, 341, 342, 343, 344, 345, 346, 347, 348, 349, 350, 351, 352, 353, 354, 355, 356, 357, 358, 359, 360, 361, 362, 363, 364, 365, 366, 367, 368, 369, 370, 371, 372, 373, 374, 375, 376, 377, 378, 379, 380, 381, 382, 383, 384, 385, 386, 387, 388, 389, 390, 391, 392, 393, 394, 395, 396, 397, 398, 399, 400, 401, 402, 403, 404, 405, 406, 407, 408, 409, 410, 411, 412, 413, 414, 415, 416, 417, 418, 419, 420, 421, 422, 423, 424, 425, 426, 427, 428, 429, 430, 431, 432, 433, 434, 435, 436, 437, 438, 439, 440, 441, 442, 443, 444, 445, 446, 447, 448, 449, 450, 451, 452, 453, 454, 455, 456, 457, 458, 459, 460, 461, 462, 463, 464, 465, 466, 467, 468, 469, 470, 471, 472, 473, 474, 475, 476, 477, 478, 479, 480, 481, 482, 483, 484, 485, 486, 487, 488, 489, 490, 491, 492, 493, 494, 495, 496, 497, 498, 499, 500, 501, 502, 503, 504, 505, 506, 507, 508, 509, 510, 511, 512, 513, 514, 515, 516, 517, 518, 519, 520, 521, 522, 523, 524, 525, 526, 527, 528, 529, 530, 531, 532, 533, 534, 535, 536, 537, 538, 539, 540, 541, 542, 543, 544, 545, 546, 547, 548, 549, 550, 551, 552, 553, 554, 555, 556, 557, 558, 559, 560, 561, 562, 563, 564, 565, 566, 567, 568, 569, 570, 571, 572, 573, 574, 575, 576, 577, 578, 579, 580, 581, 582, 583, 584, 585, 586, 587, 588, 589, 590, 591, 592, 593, 594, 595, 596, 597, 598, 599, 600, 601, 602, 603, 604, 605, 606, 607, 608, 609, 610, 611, 612, 613, 614, 615, 616, 617, 618, 619, 620, 621, 622, 623, 624, 625, 626, 627, 628, 629, 630, 631, 632, 633, 634, 635, 636, 637, 638, 639, 640, 641, 642, 643, 644, 645, 646, 647, 648, 649, 650, 651, 652, 653, 654, 655, 656, 657, 658, 659, 660, 661, 662, 663, 664, 665, 666, 667, 668, 669, 670, 671, 672, 673, 674, 675, 676, 677, 678, 679, 680, 681, 682, 683, 684, 685, 686, 687, 688, 689, 690, 691, 692, 693, 694, 695, 696, 697, 698, 699, 700, 701, 702, 703, 704, 705, 706, 707, 708, 709, 710, 711, 712, 713, 714, 715, 716, 717, 718, 719, 720, 721, 722, 723, 724, 725, 726, 727, 728, 729, 730, 731, 732, 733, 734, 735, 736, 737, 738, 739, 740, 741, 742, 743, 744, 745, 746, 747, 748, 749, 750, 751, 752, 753, 754, 755, 756, 757, 758, 759, 760, 761, 762, 763, 764, 765, 766, 767, 768, 769, 770, 771, 772, 773, 774, 775, 776, 777, 778, 779, 780, 781, 782, 783, 784, 785, 786, 787, 788, 789, 790, 791, 792, 793, 794, 795, 796, 797, 798, 799, 800, 801, 802, 803, 804, 805, 806, 807, 808, 809, 810, 811, 812, 813, 814, 815, 816, 817, 818, 819, 820, 821, 822, 823, 824, 825, 826, 827, 828, 829, 830, 831, 832, 833, 834, 835, 836, 837, 838, 8

XX
BT
H3301b0d6 P
C01
vt
E:01:01:01
T03
E:01
E
E:01:01:01
E
C
C
C

Yamamoto RT, Xu HH;

DR WPI; 2001-611495/70.

Now set: $\text{set } 1000000000$

PT antibiotics, comprise sequences of antisense nucleic acids -

PS Example 3; Seq ID No 10944; 511pp; English
v

CC The invention relates to antisense inhibitors of genes essential to

CC genes, their use in the discovery of novel antibiotics, the essential

CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The

CC invention is also useful for the identification of potential new targets for antibiotic development. The antibiotic targets are also

CC to identify proteins used in proliferation, to express these proteins and to obtain antibodies capable of binding to the targeted proteins.

CC The proteins can be used to screen compounds in rational drug discovery programmes. The antigenic acid sequence is also useful to screen

CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence represents an

CC essential prokaryotic cellular proliferation protein
CC Note: The sequence data for this patent did not form

of the printed specification, but was obtained in electronic format directly from WIPD at

cc http://www.wipo.int/pat/pct_sequences.
xx

Sequence	338 AA;
50	1
100	2
150	3
200	4
250	5
300	6
350	7
400	8
450	9
500	10
550	11
600	12
650	13
700	14
750	15
800	16
850	17
900	18
950	19
1000	20
1050	21
1100	22
1150	23
1200	24
1250	25
1300	26
1350	27
1400	28
1450	29
1500	30
1550	31
1600	32
1650	33
1700	34
1750	35
1800	36
1850	37
1900	38
1950	39
2000	40
2050	41
2100	42
2150	43
2200	44
2250	45
2300	46
2350	47
2400	48
2450	49
2500	50
2550	51
2600	52
2650	53
2700	54
2750	55
2800	56
2850	57
2900	58
2950	59
3000	60
3050	61
3100	62
3150	63
3200	64
3250	65
3300	66
3350	67
3400	68
3450	69
3500	70
3550	71
3600	72
3650	73
3700	74
3750	75
3800	76
3850	77
3900	78
3950	79
4000	80
4050	81
4100	82
4150	83
4200	84
4250	85
4300	86
4350	87
4400	88
4450	89
4500	90
4550	91
4600	92
4650	93
4700	94
4750	95
4800	96
4850	97
4900	98
4950	99
5000	100

```
Query Match      10.7%; Score 206; DB 22; Length 338;
```

Query Match	9.04;	Score 173;	DB 23;	Length 328
Best Local Similarity	28.84;	Pred. No. 2.8e-07;		

CC The present invention provides a number of nucleotide and protein
CC sequences from the *Corynebacterium* bacterium *Corynebacterium glutamicum*. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of *Corynebacterium* bacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from *Corynebacterium* bacterium, and identifying a homologue of a gene derived
CC from *Corynebacterium* bacterium. *Corynebacterium* bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a protein described
CC in the exemplification of the invention.

CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC European Patent Office.

XX Sequence 337 AA;

Query Match 8.8%; Score 168; DB 22; Length 337;
 Best Local Similarity 29.4%; Pred. No. 8.2e-07;
 Matches 69; Conservative 31; Mismatches 93; Indels 42; Gaps 11;

QY 2 SVPTTKAVIIEGDAVVKTDVSVPELKEGTALVKEAVAGNPTMKHAIYKGE--GS 59
 DB 20 SIPSDDPFLID-----VEVDRPTPGPHDLVHIEAVSINPVDTK-VRMAGKQKHPR 71
 QY 60 ILGCDIAGTVVVLGNASTDLKVGDTGFGFVHGASQOTDKNGAFAYARVVPLEPKSNL 119
 DB 72 ILGFDAGAEVAVAGSQV-TLFNVGDK--VFYAGSNQ---RPSGNAY----- 112

120 THSTADE--ISEGPVK-NFESAASLPVSLTTAGVSLCHNLGSKMEWHSPSTPQHTPLLIW 176
 113 --QVDERLVGHAPQSLGAHDAALPLVALTAMESLFDRLG-----VTQSTGTGLVL 163
 QY 177 GGATAVGQOLIQVAHINAYTKIVYASKKHEKLKSYGADVPFYHDAGVIEOI 231
 DB 164 GSGGVPSPALIQARALTGKLVATASRPESQEWYTKLGAHEVID-HSKDLSBOI 217

RESULT 14

AA879594
 ID AAB79594 standard; Protein; 337 AA.

AC AAB79594;

XX 30-APR-2001 (first entry)

DE Corynebacterium glutamicum SMP protein sequence SEQ ID NO:704.

KM Corynebacterium glutamicum; carbon metabolism and energy production;
 KM SMP protein; sugar metabolism and oxidative phosphorylation protein;
 KM fine chemical production; organic acid; proteinogenic amino acid;
 KM nonproteinogenic amino acid; purine base; pyrimidine base; nucleoside;
 KM nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;
 KM carbohydrate; aromatic compound; vitamin; cofactor; polypeptide; enzyme;
 KM diagnosis; Corynebacterium diptheriae; evolutionary study.

XX Corynebacterium glutamicum.

WO200100844-A2.

04-JAN-2001.

XX 23-JUN-2000; 2000WO-1B00943.

XX 25-JUN-1999; 99US-0141031.

PR 08-JUL-1999; 99DE-1031412.

PR 08-JUL-1999; 99DE-1031413.

PR 08-JUL-1999; 99DE-1031419.

PR 08-JUL-1999; 99DE-1031420.

PR 08-JUL-1999; 99DE-1031424.

PR 08-JUL-1999; 99DE-1031431.

PR 08-JUL-1999; 99DE-1031433.

PR 08-JUL-1999; 99DE-1031434.

PR 08-JUL-1999; 99DE-1031510.

PR 08-JUL-1999; 99DE-1031562.

PR 08-JUL-1999; 99DE-1031634.

PR 09-JUL-1999; 99DE-1032180.

PR 09-JUL-1999; 99DE-1032227.

PR 09-JUL-1999; 99DE-1032230.

PR 09-JUL-1999; 99US-0143208.

PR 14-JUL-1999; 99DE-1032524.

PR 14-JUL-1999; 99DE-1032573.

PR 14-JUL-1999; 99DE-1033005.

PR 27-AUG-1999; 99DE-1040765.
 PR 31-AUG-1999; 99US-0151572.
 PR 03-SEP-1999; 99DE-1042076.
 PR 03-SEP-1999; 99DE-1042079.
 PR 03-SEP-1999; 99DE-1042086.
 PR 03-SEP-1999; 99DE-1042087.
 PR 03-SEP-1999; 99DE-1042088.
 PR 03-SEP-1999; 99DE-1042095.
 PR 03-SEP-1999; 99DE-1042123.
 PR 03-SEP-1999; 99DE-1042125.

PA (BADI) BASF AG.

PI Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberbauer G;

XX WPI, 2001-061975/07.

DR N-PSDB; AAF71711.

XX

XX

PT New isolated Corynebacterium glutamicum nucleic acid encoding a sugar

PT metabolism and oxidative phosphorylation protein for production or

PT modulation of production of fine chemicals e.g. amino acids,

PT carbohydrates or enzymes -

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Sequence 337 AA;

Query Match 8.8%; Score 168; DB 22; Length 337;
 Best Local Similarity 29.4%; Pred. No. 8.2e-07;
 Matches 69; Conservative 31; Mismatches 93; Indels 42; Gaps 11;

QY 2 SVPTTKAVIIEGDAVVKTDVSVPELKEGTALVKEAVAGNPTMKHAIYKGE--GS 59
 DB 20 SIPSDDPFLID-----VEVDRPTPGPHDLVHIEAVSINPVDTK-VRMAGKQKHPR 71

QY 60 ILGCDIAGTVVVLGNASTDLKVGDTGFGFVHGASQOTDKNGAFAYARVVPLEPKSNL 119
 DB 72 ILGFDAGAEVAVAGSQV-TLFNVGDK--VFYAGSNQ---RPSGNAY----- 112

QY 120 THSTADE--ISEGPVK-NFESAASLPVSLTTAGVSLCHNLGSKMEWHSPSTPQHTPLLIW 176
 DB 113 --QVDERLVGHAPQSLGAHDAALPLVALTAMESLFDRLG-----VTQSTGTGLVL 163

QY 177 GGATAVGQOLIQVAHINAYTKIVYASKKHEKLKSYGADVPFYHDAGVIEOI 231
 DB 164 GSGGVPSPALIQARALTGKLVATASRPESQEWYTKLGAHEVID-HSKDLSBOI 217

RESULT 15

AA828581
 ID AAG28581 standard; Protein; 363 AA.

AC AAG28581;

XX

PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 8.7%; Score 167; DB 21; Length 363;
Best Local Similarity 26.1%; Pred. No. 1.1e-06;
Matches 72; Conservative 47; Mismatches 91; Indels 66; Gaps 15;

QY 2 SVPTTOKAVIIE--GDKAV--VKTDVSPBELKGTALVKEVAAGNPTDMK----- 48
DB 24 SVFTGCRVILPRFGGPEVLEFLENVFNLNENEVLPKAAVSVPFLCRLRAGYGRSV 83
QY 49 ---HIAVKIGPESGILGDIAGTVVVLGPNASTDLKVGDTGFGFVHGASQTDPKNGAFAB 105
DB 84 PQPHLP-----IVGRDVSGEVAIAIGSVKS-LKVGGEVFGALHPTA-----LRGTYYD 131
QY 106 YARVVPPLFYKSNLHSTADEISEGP--VKNFESAASLPVSLTTAGVSLCHHLGSKMEWH 163
DB 132 YGIL-----SEDELTEKPSISISHE--ASAIFFAALTAWRAL----- 166
QY 164 PSTPQHT--HPTLIMGATAVGQQLIOVKHINATKIYTVASKKHKLKSGYADVD 221
DB 167 KSNARTTEGRLVFGGGLVFGGSAIQLVASGCHVTASCVGQTK--DRILAA-GAQAVD 224
QY 222 YHDAGVTEQIKSKYPMLOHVIVAVGSEDSIPEAYKV 257

DB 225 YTTEDIEHAVKGF---DAVLDITGG-----PETERI 253

Search completed: June 17, 2003, 10:48:46
Job time : 44 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OW protein - protein search, using sw model

Run on: June 17, 2003, 10:44:38 ; Search time 19 Seconds
(without alignments)
572.973 Million cell updates/sec

Title: US-10-081-644-2

Perfect score: 1920

Sequence: 1 MSVPTTQKAVIIIEGDKAVK.....ALTEGKKGKKNKVVAVRL 370

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA:*
1: /cgn2_6/prodata/1/1aa/5A_COMB.pep:*
2: /cgn2_6/prodata/1/1aa/5B_COMB.pep:*
3: /cgn2_6/prodata/1/1aa/6A_COMB.pep:*
4: /cgn2_6/prodata/1/1aa/6B_COMB.pep:*
5: /cgn2_6/prodata/1/1aa/6CTOS_COMB.pep:*
6: /cgn2_6/prodata/1/1aa/backfillseq.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	266.5	13.9	363	4	US-09-215-694-5 Sequence 5, Appl1
2	158.5	8.3	396	4	US-08-860-656B-2 Sequence 2, Appl1
3	158.5	8.3	396	4	US-08-860-656B-3 Sequence 3, Appl1
4	153.5	8.0	336	1	US-08-713-254-2 Sequence 2, Appl1
5	153.5	8.0	336	4	US-08-855-767-2 Sequence 325B, Ap
6	151.5	7.9	340	4	US-09-134-001C-3258 Sequence 81, Appl1
7	149.5	7.8	322	4	US-09-154-750A-81 Sequence 4, Appl1
8	131.5	6.8	396	4	US-08-860-656B-4 Sequence 10, Appl1
9	126	6.6	2532	4	US-09-215-694-10 Sequence 94, Appl1
10	121	6.3	345	2	US-07-857-224B-94 Sequence 93, Appl1
11	118	6.1	344	2	US-07-857-224B-93 Sequence 3, Appl1
12	115.5	6.0	2152	4	US-09-036-987A-3 Sequence 10, Appl1
13	115.5	6.0	2152	4	US-09-370-700-3 Sequence 4, Appl1
14	114	5.9	3724	2	US-08-804-198-4 Sequence 92, Appl1
15	114	5.9	3724	2	US-08-804-198-4 Sequence 95, Appl1
16	111	5.8	300	2	US-08-828-010-2 Sequence 2, Appl1
17	108	5.6	344	2	US-07-857-224B-92 Sequence 4, Appl1
18	108	5.6	349	2	US-07-857-224B-95 Sequence 4, Appl1
19	108	5.6	3567	2	US-07-642-734C-4 Sequence 4, Appl1
20	107.5	5.6	3567	2	US-08-439-009A-4 Sequence 3, Appl1
21	107.5	5.5	3816	4	US-09-428-517-3 Sequence 3374, Ap
22	105.5	5.5	354	4	US-09-134-001C-3374 Sequence 216, App
23	105.5	5.5	1007	4	US-08-961-083-216 Sequence 2, Appl1
24	105	5.5	376	1	US-08-608-241-2 Sequence 2, Appl1
25	105	5.5	376	2	US-08-822-182-2 Sequence 2, Appl1
26	105	5.5	376	2	US-08-919-953-2 Sequence 2, Appl1
27	105	5.5	376	4	US-09-192-983-2 Sequence 2, Appl1

28	102.5	5.3	346	2	US-07-857-224B-96 Sequence 96, Appl1
29	101	5.3	352	2	US-08-805-191-2 Sequence 2, Appl1
30	100.5	5.2	461	4	US-09-122-210-2 Sequence 2, Appl1
31	100.5	5.2	461	4	US-09-443-681-2 Sequence 89, Appl1
32	99.5	5.2	324	4	US-09-154-750A-89 Sequence 2, Appl1
33	98.5	5.1	3739	4	US-09-320-878-2 Sequence 33, Appl1
34	98.5	5.1	3739	4	US-09-105-537-6 Sequence 6, Appl1
35	98.5	5.1	11877	4	US-09-105-537-6 Sequence 5, Appl1
36	96.5	5.0	7257	3	US-09-335-409-5 Sequence 5, Appl1
37	96.5	5.0	7257	4	US-09-568-102-5 Sequence 5, Appl1
38	96.5	5.0	7257	4	US-09-567-969-5 Sequence 5, Appl1
39	96.5	5.0	7257	4	US-09-568-480-5 Sequence 5, Appl1
40	96.5	5.0	7257	4	US-09-568-486-5 Sequence 5, Appl1
41	96.5	5.0	7257	4	US-09-568-472-5 Sequence 5, Appl1
42	96.5	5.0	7257	4	US-09-567-899-5 Sequence 7, Appl1
43	96	5.0	1098	1	US-07-777-715-7 Sequence 2, Appl1
44	96	5.0	1098	1	US-08-170-126-2 Sequence 2, Appl1
45	96	5.0	1098	3	US-08-954-418-2 Sequence 2, Appl1

ALIGNMENTS

RESULT 1					
US-09-215-694-5					
; Sequence 5, Application US/09215694B					
; Patent No. 6391583					
; GENERAL INFORMATION:					
; APPLICANT: Wisconsin Alumni Research Foundation					
; APPLICANT: Hutchinson, Charles R.					
; APPLICANT: Kennedy, Jonathan n.m.i					
; APPLICANT: Park, Cheonsuek n.m.i					
; TITLE OF INVENTION: METHOD OF PRODUCING ANTIHYPERCHOLESTEROLEMIC AGENTS					
; FILE REFERENCE: 960296.95718					
; CURRENT APPLICATION NUMBER: US/09/215,694B					
; CURRENT FILING DATE: 1999-12-18					
; NUMBER OF SEQ ID NOS: 36					
; SOFTWARE: PatentIn Ver. 2.0					
; SEQ ID NO 5					
; LENGTH: 363					
; TYPE: PRT					
; ORGANISM: Aspergillus terreus					
US-09-215-694-5					
Query Match					
Best Local Similarity 26.3%; Pred. No. 9.4e-19;					
Matches 103; Conservative 69; Mismatch 156; Indels 63; Gaps 17;					
QY	4	PTTQKAVII-EGDKAVKTDVSVPELKEGTALVKEVAAGNPTDKHIAKIGPGSILG	62		
DB	9	PTQQTALVNDHDEVTVMNAACFPLPRDQYVREVAVALNPSDK-MRGQPATWAFUG	67		
QY	63	CDIAGTVKLGPNASTDKVGTGFGFVHGASQTPKNGAFAY-ARVYPLFYKSN	118		
DB	68	TVAGTVVAVGSDV-THIOGDRVYGAQNMCPRPDQCAFQYVTRGRVVA-----	119		
QY	119	LTHSTADELSSEPVKNFSSASLPVSLTTAGVSLCHHLSKMEKMPST-PQTHRP-L	173		
DB	120	-----KIPKG--LSFQMAALPAGISTAGIAM-KLGLPLP-SPSADQPTHSKPPVV	168		
QY	174	LITGATVAGQGLIOVAKHINAYKIIVVASKHEKLKSYGADVDFVDHAGVLEQIKS	233		
DB	169	LVYGSSTATATVYTMQML-LSGYIYIAT-CSFHNFDLAKSRGAEEVFDRAVNLQITRT	226		
QY	234	KYPLQHVIVDAVSGSDSIPEAYKYVATDSLPAITLLEVPMTIESIPEE-IRKDNVKIDIT	291		
DB	227	YTKNNLRYALDCITVVESTTFCEFAIGRAGHY-----VSLNPPRHAATRKQVTTDWT	280		
QY	232	LITRASGEIILGARF-----PASPEYHEATVFKVKINHLNNGDIHNNI	339		
DB	281	-----LGPITFEGSGTWPAYPGRPGSEERQFGEDLMRIAGOLVEDGRLVHHPPL	329		
QY	340	KVFSNGLDVPAITGKIGKKNKVVAVRL 370			

TELEFAX: (703) 413-2220
 TELEX: 248855 OPAT UR
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 336 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-713-254-2

Query Match 8.0%; Score 153.5; DB 1; Length 336;
 Best Local Similarity 24.8%; Pred. No. 2.7e-07;
 Matches 101; Conservative 50; Mismatches 125; Indels 131; Gaps 22;

1 MSVPTTOKAVIIIEGDAV-VKTDVSVPELKEGTALVKVEAVGNPTDMKHIYK---IGP 56
 1 MSIPSSQYGFVFNKSGSLNRLDLPVHKPKAGQLIKVDVGLCHSD-LHVIYEGIDCG- 58
 57 EGSIIIGCDIAGTVKVLGPNASTDLKVGDY-----GFGFVHGAQTDPKN----- 100
 59 DNYVNGHEIAGTVAAVGDV-INYKVGDRVACVPGNGCGCKYCRGALDNCNKAFGDW 117
 101 -----GARAEYKRVYPPLEFYKSNLTHSTADEISEGPKNFESASLPVSLTTAGVSLCH 155
 118 GLGYDGGYQOYLIVTRP-----RNLRS-IPDNVSAD-----VAAASTDAVLTP-----YH 161
 156 LGSKMEHSPSTPHTHPLIMWGATVAGQQLIQVAKHINAYKIYVASKKEH--KLKLS 213
 162 AIKMAQVSTP---SNILII--GAGGLGSMALQVAK--AFGAKTVLDKKKEARDQAK 212
 214 YGADVDPYHDAGVIEQIKSKYPNLQHVDAVGSSEDSIPEAYKVTADSLPATLLEVPM 273
 213 LGADVY----- 219
 274 IESIEEIRKDNVKIDITLLYRASQGEILLGATRPASPEYHE--ATYKVFYINPHL-- 329
 220 -ETLPESSIPGS-----FSACDFVSVQAT-FDVOCQKVEPBGVIMPGVIGAPNLSF 269
 330 NNGDHHNNIKVF-----SNGLDVPAITBEGIKSKKNVYKVARL 370
 270 NIGDLALREIRILGSPFGTNDLVDLKL---VSEGKVPVRSACL 313

RESULT 5
 US-08-855-767-2
 Sequence 2, Application US/08855767

GENERAL INFORMATION:
 APPLICANT: KOJIMA, TOMOKO
 APPLICANT: YAMAMOTO, HIROAKI
 APPLICANT: KAMADA, NAOKI
 APPLICANT: MATSUYAMA, AKINOBU
 TITLE OF INVENTION: NOVEL ENZYME, A METHOD TO PREPARE SAID
 TITLE OF INVENTION: ENZYME, A DNA SEGMENT ENCODING SAID ENZYME, A TRANSFORMANT
 TITLE OF INVENTION: CONTAINING SAID DNA SEGMENT AND A METHOD OF
 TITLE OF INVENTION: OPTICALLY ACTIVE ALCOHOL USING SAID ENZYME
 NUMBER OF SEQUENCES: 11
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MATER & NEUSTADT,
 ADDRESSEE: P.C.
 STREET: 1755 S. Jefferson Davis Highway, Suite 400
 CITY: Arlington
 STATE: Virginia
 COUNTRY: U.S.A.
 ZIP: 22202
 COMPUTER READABLE FORM:
 MEDIUM TYPE: floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/855,767
 FILING DATE:

CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/311,328
 FILING DATE:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: JP 5-337191
 FILING DATE: 28-DEC-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: JP 6-181308
 FILING DATE: 02-AUG-1994
 ATTORNEY/AGENT INFORMATION:
 NAME: Oblon, No. 6255092man F.
 REGISTRATION NUMBER: 24,618
 REFERENCE/DOCKET NUMBER: 6423-001-0
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (703) 413-3000
 TELEFAX: (703) 413-2220
 TELEX: 248855 OPAT UR
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 336 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-855-767-2

Query Match 8.0%; Score 153.5; DB 4; Length 336;
 Best Local Similarity 24.8%; Pred. No. 2.7e-07;
 Matches 101; Conservative 50; Mismatches 125; Indels 131; Gaps 22;

1 MSVPTTOKAVIIIEGDAV-VKTDVSVPELKEGTALVKVEAVGNPTDMKHIYK---IGP 56
 1 MSIPSSQYGFVFNKSGSLNRLDLPVHKPKAGQLIKVDVGLCHSD-LHVIYEGIDCG- 58
 57 EGSIIIGCDIAGTVKVLGPNASTDLKVGDY-----GFGFVHGAQTDPKN----- 100
 59 DNYVNGHEIAGTVAAVGDV-INYKVGDRVACVPGNGCGCKYCRGALDNCNKAFGDW 117
 101 -----GARAEYKRVYPPLEFYKSNLTHSTADEISEGPKNFESASLPVSLTTAGVSLCH 155
 118 GLGYDGGYQOYLIVTRP-----RNLRS-IPDNVSAD-----VAAASTDAVLTP-----YH 161
 156 LGSKMEHSPSTPHTHPLIMWGATVAGQQLIQVAKHINAYKIYVASKKEH--KLKLS 213
 162 AIKMAQVSTP---SNILII--GAGGLGSMALQVAK--AFGAKTVLDKKKEARDQAK 212
 214 YGADVDPYHDAGVIEQIKSKYPNLQHVDAVGSSEDSIPEAYKVTADSLPATLLEVPM 273
 213 LGADVY----- 219
 274 IESIEEIRKDNVKIDITLLYRASQGEILLGATRPASPEYHE--ATYKVFYINPHL-- 329
 220 -ETLPESSIPGS-----FSACDFVSVQAT-FDVOCQKVEPBGVIMPGVIGAPNLSF 269
 330 NNGDHHNNIKVF-----SNGLDVPAITBEGIKSKKNVYKVARL 370
 270 NIGDLALREIRILGSPFGTNDLVDLKL---VSEGKVPVRSACL 313

RESULT 6
 US-09-134-001C-3258
 Sequence 3258, Application US/09134001C

GENERAL INFORMATION:
 APPLICANT: Lynn Doucette-Stamm et al
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
 TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
 FILE REFERENCE: GTC-007
 CURRENT APPLICATION NUMBER: US/09/134,001C
 CURRENT FILING DATE: 1998-08-13
 PRIOR APPLICATION NUMBER: US 60/064,964
 PRIOR FILING DATE: 1997-11-08
 PRIOR APPLICATION NUMBER: US 60/055,779

PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 3258
LENGTH: 340
TYPE: PR
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3258

Query Match 7.8%; Score 151.5; DB 4; Length 340;
Best Local Similarity 27.8%; Pred. No. 4.4e-07;
Matches 67; Conservative 33; Mismatches 102; Indels 39; Gaps 11;

QY 6 TQAAVIEGDKAVKTD-----VSVELKRGALVVEAVAGNPTDKHIAIKIGP 56
DB 4 TMAAI---GFKSSFOLEDEGNCFEFNFEDIPHSGHEILVVKQISIVPVDTKQRTMPVDK 60
QY 57 EGSIIIGCDIAGTVKVGPNASTDLKVGDTGFGFVHAGSQTDPKNGAPAEVARYPPFYK 116
DB 61 APRVLGPDVAVIEKIGDOVSM--FQEGDVVF-----YSGSPNONGSNEETQLEEYL--- 111
QY 117 SNLTHSTADRISEGPVK-NFESAASLPVSLTTAGVSLCHHLSKMEWHPSTPOHTHPLLI 175
DB 112 -----VAKPTLTKSEQAASLPGLTGLTAVETLFDVFGISKE--PS-ENKKSLLI 158
QY 176 WSGATVAGQQLIVAKHINAYTKIVVASKKH-KLIKSYGADVDPDYDAGVIEQIKSK 234
DB 159 INGAGVGSIATQIAKPYG--LKVITIASREDITIKSVNMAGDVVLN-HKKDLSCQFKN 215
QY 235 Y 235
DB 216 H 216

RESULT 7

US-09-154-750A-81
Sequence 81, Application US/09154750A
Patent No. 6432640
GENERAL INFORMATION:
APPLICANT: Vogelstein, Bert
APPLICANT: Kinzler, Kenneth
APPLICANT: Polyak, Kornelia
TITLE OF INVENTION: p53-Induced Apoptosis
FILE REFERENCE: 1107.75357
CURRENT APPLICATION NUMBER: US/09/154.750A
PRIOR FILING DATE: 1998-09-17
PRIOR APPLICATION NUMBER: 60/059,153
PRIOR FILING DATE: 1997-09-17
PRIOR APPLICATION NUMBER: 60/079817
PRIOR FILING DATE: 1998-03-30
NUMBER OF SEQ ID NOS: 93
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 81
LENGTH: 322
TYPE: PR
ORGANISM: Homo sapiens
US-09-154-750A-81

Query Match 7.8%; Score 149.5; DB 4; Length 322;
Best Local Similarity 26.9%; Pred. No. 6.4e-07;
Matches 64; Conservative 34; Mismatches 105; Indels 35; Gaps 9;

QY 14 GDAVAVTDSVPELKEGTALVKEAVAGNPTDKHIAIKIGP---EGSIIIGCDIAGTV 70
DB 11 GREPLVYKEVAKSPGEGEVLLTKVAAASALNRADLMKQCGDYDPPPGASNLIGLEASGHVA 70
QY 71 KLGPMASDTLKGNDTGGFPHGASQDTPKNGAPAEVARYPPFYKSNLHSTADEISG 130
DB 71 ELGGCGGHWKIGDTAMALLPGGQ-----AQYVTEBGLM-----PIPEG 112
QY 131 PVKMFESAASLPVSLTTAGVSLCHHLSKMEWHPSTPOHTHPLLIWAGATVAGQQLIOVA 190
DB 113 --LTVIQAIAIPKMLTH-FQLHLVGN-----VQADYVLIHAGLSGVGAIAQLT 161

QY 191 KHNATKIVTVASKKHEKLSYGADVDYDAGVIE-QIK-SKYPNLQHVIVANG 246
DB 162 RWAGA-IPLVTAGSQKQLQMAEKLGAAAGFNKEDFSEATLTKTGAGVNLILDCIG 218

RESULT 8

US-08-860-656B-4
Sequence 4, Application US/08860656B
Patent No. 6297028
GENERAL INFORMATION:
APPLICANT: Taniguchi, Tadatsugu
APPLICANT: Shibuya, Hiroshi
APPLICANT: Barsumian, Edward L.
TITLE OF INVENTION: IL-2R-Associated Polypeptides and DNA Molecules Coding Therefor
FILE REFERENCE: 0652.1640000
CURRENT APPLICATION NUMBER: US/08/860,656B
PRIOR FILING DATE: 1998-01-15
PRIOR APPLICATION NUMBER: PCT/EP95/05123
PRIOR FILING DATE: 1995-12-23
NUMBER OF SEQ ID NOS: 18
SOFTWARE: PatentIn version 3.0
SEQ ID NO 4
LENGTH: 396
TYPE: PR
ORGANISM: Mouse p43
US-08-860-656B-4

Query Match 6.8%; Score 131.5; DB 4; Length 396;
Best Local Similarity 27.6%; Pred. No. 6.1e-05;
Matches 67; Conservative 31; Mismatches 76; Indels 69; Gaps 13;

QY 34 LVKEVAVAGNPTDKM-----HIAVIGPESGI-LGCDIAGTVKLGPN 75
DB 75 IIKVHAASVNPIDVNMRSYGATATLNMKSDPLHMKTK-GEFPLTLGRDVSQVMECG-- 131
QY 76 ASDTLKV---CPTGREGFVHAGSQTDPKNGAPAEVARYPPFYKSNLHSTADEISG 132
DB 132 --LDVRYFQPGDEWV-----AAVPWKQGLTFEYVV-----SGNEVSHKEK 171
QY 133 K-NFESAASLP-VSLT-----TAGVSLCHHLSKMEWHPSTPOHTHPLLIWAGATVAG 183
DB 172 SLTHQAASLPYVALTAMAINKVGSLDRNCKGR-----ALLIGASGVG 218
QY 184 QQLIQVAKHINAYTKIVVASKKHLSYGADVDYDAGVIEQIKSKYPNLQHVIV 243
DB 219 TFAIQVMKAMGAH--VTAVCSKDASELYRKLGADEVIVDTLGSVEQLKS-LKLCAFIID 275
QY 244 AVG 246
DB 276 NVG 278

RESULT 9

US-09-215-694-10
Sequence 10, Application US/09215694B
Patent No. 6391583
GENERAL INFORMATION:
APPLICANT: Wisconsin Alumni Research Foundation
APPLICANT: Hutchinson, Charles R.
APPLICANT: Kennedy, Jonathan n.m.i.
APPLICANT: Park, Cheonsuck n.m.i.
TITLE OF INVENTION: METHOD OF PRODUCING ANTIHYPERCHOLESTEROLEMIC AGENTS
FILE REFERENCE: 960296.95718
CURRENT APPLICATION NUMBER: US/09/215,694B
PRIOR FILING DATE: 1999-12-18
NUMBER OF SEQ ID NOS: 36
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 10
LENGTH: 2532
TYPE: PR
ORGANISM: Aspergillus terreus
US-09-215-694-10

Query Match 6.6%; Score 126; DB 4; Length 2532;
Best Local Similarity 23.4%; Pred. No. 0.0043;
Matches 56; Conservative 39; Mismatches 92; Indels 52; Gaps 10;

QY 18 VKKIDVSEPELKEGRLVYK--VEAVAGN--PTDW-----KAIAYIGP-EGSI 60
DB 1816 LRLRLIOTFGLDLSLHPTKRNVDYEPDLDPDWVEIERRAFGLNFRDIMVAMGOLSENV 1875
QY 61 LGGDIAGTVKLGPNAST---DLKVGDTGFGVHGASOTDPKNGAPAEVARYPPPLFYKS 117
DB 1876 MPEECAGVTSLETRRTTAPGLAVGDRVCALMG-----HMS 1914
QY 118 NLTHSTADEISEGPVKNFESAASLPVSLTTAGVSLCHLGSXQEMHPSPTQHTPLIWG 177
DB 1915 RTTSTRNVRIPEPLISFPHASIPLAFTTAYISL--YTVARI-----LPGFT--VLTHA 1965
QY 178 GATVAGQOLIQAQHINATYKIVTVASKGHEKLKSY--GADVDPDYHDAGVIEQISK 234
DB 1966 GAGGVQQAIIIAQLTGAEVFTTAGSETKRNLLIDKFLHDPDHVFSRDSFVDGIKTR 2024

RESULT 10
US-07-857-224B-94
Sequence 94, Application US/07857224B
Patent No. 5958784
GENERAL INFORMATION:
APPLICANT: Benner, Steven A.
TITLE OF INVENTION: Predicting Folded Structures of Proteins
NUMBER OF SEQUENCES: 114
CORRESPONDENCE ADDRESS:
ADDRESSEE: Steven A. Benner
STREET: Hadlaubstrasse 151
CITY: Zurich
STATE: none
COUNTRY: Switzerland
ZIP: (note: this is an international post code) CH-8092
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch diskette, 1.4 Mb storage
COMPUTER: Apple Macintosh
OPERATING SYSTEM: Macintosh 7.0
SOFTWARE: Microsoft Word
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/857,224B
FILING DATE: 03/25/92
CLASSIFICATION: 436
PRIOR APPLICATION DATA: none
TELECOMMUNICATION INFORMATION:
TELEPHONE: (International) 41 1 632 2830
TELEFAX: (International) 41 1 262 2437
TELEX: none
INFORMATION FOR SEQ ID NO: 94:
SEQUENCE CHARACTERISTICS:
LENGTH: 345
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Saccharomyces cerevisiae
FEATURE: Alcohol dehydrogenase, Table 3 Column 17
PUBLICATION INFORMATION:
AUTHORS:
AUTHORS: Joernvall, H.
AUTHORS: Pettersson, M.
AUTHORS: Jeffery, J.
TITLE: Alcohol dehydrogenases
JOURNAL: Proceedings of the National Academy of Sciences, USA
VOLUME: 78
PAGES: 4226-4230
DATE: 1981
US-07-857-224B-94

Query Match 6.3%; Score 121; DB 2; Length 345;

Best Local Similarity 22.1%; Pred. No. 0.00057;
Matches 88; Conservative 52; Mismatches 142; Indels 116; Gaps 21;

QY 2 SVPTQKAVIIEGDAVVK-IDVSEPELKEGRLVYKVAVAGNP-----DKHIAVYK 53
DB 2 AIPKQKQVIFYENKGLKXKIDIPVKKPKEILLNMKYSGVCHTDLHAMHGWDP-LPVK 60
QY 54 I---GPGSILIGCDIAGTVVYKLGPNASTDLKVGQ--TSGFVHGASOT-----96
DB 61 LPVGHGEG-----AGVYKLGSNVYK-GMKVGLAGIKMLNGSCMTCEPESGHSNC 112
QY 97 -----DPKGAFAEYARYVPPLEFYKSNLTHSTADEISEGPVKNFESAASLPVSLTTAG 149
DB 113 PDADLSGYTHDSFGQFA-----TDAIQAAKIQCGTDLAEV-APILCG 156
QY 150 VSLCHHLG---SKEMHPSPTQHTPLIINGGATVAGQOLIQAQHINATYKIVTVASK 205
DB 157 VTVYKALKEADLKAGDW-----VAISGAGLGSIAVQYATVANGYRVLGIDAGE 206
QY 206 KHEKLLKSYGADVDPDYHDAGVIEQISKYVNLGHVIDAVGSEDSIPEAYKVTADSLPAT 265
DB 207 EKER--KVLGGEVFTDF-----TKTKM-----VSDIQEATK---GGRHG 241
QY 266 LLEVPMPTIESIPEIRKDNVKIDITLLY-RASGOEILGATRPAS-----PEYHEATVYK 320
DB 242 VINV-----GVSE-----AAISLSTEVYRCGVVVVG---LPANAYVSEVSHVYK 286
QY 321 FVKFIPHLNNGDIIHMANIKVFSNGLDVPPLTBSQIKE 358
DB 287 SINIVSYGNRADTREALDFPSRLIKSPKIKYGLSE 324

RESULT 11
US-07-857-224B-93
Sequence 93, Application US/07857224B
Patent No. 5958784
GENERAL INFORMATION:
APPLICANT: Benner, Steven A.
TITLE OF INVENTION: Predicting Folded Structures of Proteins
NUMBER OF SEQUENCES: 114
CORRESPONDENCE ADDRESS:
ADDRESSEE: Steven A. Benner
STREET: Hadlaubstrasse 151
CITY: Zurich
STATE: none
COUNTRY: Switzerland
ZIP: (note: this is an international post code) CH-8092
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch diskette, 1.4 Mb storage
COMPUTER: Apple Macintosh
OPERATING SYSTEM: Macintosh 7.0
SOFTWARE: Microsoft Word
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/857,224B
FILING DATE: 03/25/92
CLASSIFICATION: 436
PRIOR APPLICATION DATA: none
TELECOMMUNICATION INFORMATION:
TELEPHONE: (International) 41 1 632 2830
TELEFAX: (International) 41 1 262 2437
TELEX: none
INFORMATION FOR SEQ ID NO: 93:
SEQUENCE CHARACTERISTICS:
LENGTH: 344
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Aspergillus
FEATURE: Alcohol dehydrogenase, Table 3 Column 16
PUBLICATION INFORMATION:
AUTHORS:

AUTHORS: Joernvall, H.
 AUTHORS: Peterson, M.
 AUTHORS: Jeffery, J.
 TITLE: Alcohol dehydrogenases
 JOURNAL: Proceedings of the National Academy of Sciences, USA
 VOLUME: 78
 PAGES: 4226-4230
 DATE: 1981
 US-07-857-2248-93

Query Match 6.1%; Score 118; DB 2; Length 344;
 Best Local Similarity 21.0%; Pred. No. 0.0011;
 Matches 86; Conservative 52; Mismatches 117; Indels 124; Gaps 19;

QY 2 SVPTTQKAVII-EGDKAVVTVDSVPELKEGTALVKEVAAGNPT-----DKKIIAYK 53
 1 SIPTQKAVIIFESNGLEHKDIPVFKPKNELLINVKSGVCHTDLHAWGWP-LPTK 59
 54 I-----GPESSILGCDIAGTVVKLGPNASTDLKVD-TGFGFVHGASQT----- 96
 60 LPLVGHGEG-----AGVVVGMEENK-GMKIGDYGIKWLNGSCWACEYCELGESNC 111
 QY 97 -----DPKNGAFAEYARYVPLFYKSNLTHSTADEISEGVPYNFESASLPVSLTTAG 149
 112 PHADLSGYTHDGSFOETA-----TADAV-----QAAHIPOGTDLAE 147
 QY 150 VS--LCHHLGSKMEHPSTPOHTHPLIMGATAVGOOLIOVAKHINAYTKI-VTVASKK 206
 148 VAPILGAGITVYALKASNLKAGHMAISGAGLGSIAVQAKAMGYTVVLGIDGDPK 207
 QY 207 HEKLLSYGADVDVFDYHDAGVIEQIKSKYENLGHVIDAVSSEDSIPAYVTVADSLPATL 266
 208 EETSL-----GGEVF-----IDFKKQIVSAVVAIVGAGHII 242
 DB 267 -LEVPEWTEISIEPEIRKONVKIDITLLYRASGOEILLGATRPFA-----SPEYHEATVKF 321
 243 NVSESEALIA-----STRYCAANGVAVVG--LPAKAKCSDPVFNHYKS 286
 QY 322 VKSINPLNNGDIHNNIKVFSN-----GLDDVPALTGECIKEGK 360
 DB 287 ISIVSGVGRADTREALDFPARGLVKSPIKVGLSLPEIYEMERKQ 335

RESULT 12
 US-09-036-987A-3

Sequence 3, Application US/09036987A
 Patent No. 6143526

GENERAL INFORMATION:

APPLICANT: Baltz, Richard H.
 APPLICANT: Broughton, Mary C.
 APPLICANT: Crawford, Kathryn P.
 APPLICANT: Madduri, Krishnamurthy
 APPLICANT: Merlo, Donald J.
 APPLICANT: Treadway, Patti J.
 APPLICANT: Turner, Jan R.
 APPLICANT: Waldron, Clive
 TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
 NUMBER OF SEQUENCES: 39
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Dow Agrosciences LLC Patent Department
 STREET: 9330 Zionsville Road
 CITY: Indianapolis
 STATE: Indiana
 COUNTRY: USA
 ZIP: 46268
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/036.987A

FILING DATE: 09-MAR-1998
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Stuart, Donald R
 REGISTRATION NUMBER: 28,479
 REFERENCE/DOCKET NUMBER: 50,608
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (317)337-4816
 TELEFAX: (317)337-4847
 INFORMATION FOR SEQ ID NO: 3:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2152 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-09-036-987A-3

Query Match 6.0%; Score 115.5; DB 4; Length 2152;
 Best Local Similarity 22.6%; Pred. No. 0.039;
 Matches 79; Conservative 48; Mismatches 129; Indels 93; Gaps 19;

QY 17 AVVKTVDSVPELKEGTALVKEVAAGNPTDMKHIAVKGPESSILGCDIAGTVVKLGPN 76
 1419 ALVDEPTATAPLGDSGVRIAMAGVNRD-ALINLGMVPGVASIGSEAGVVEVETGCV 1477
 QY 77 STDLYKGTGFGFVHGASQTDPRNGAFAEYARYVPLFYKSNLTHSTADEISEGVPKFE 136
 1478 -TGLAPGDRVMGVI-----PK-----AVADHRMVTIRPAG--WSPA 1515
 QY 137 SAASLPVSLTTAGVSLCHHLGSKMEHPSTPOHTHPLIMGATAVGOOLIOVAKHINAY 196
 1516 KIASVPIVFLIAYVALVDLAGR--PG-----ESLIVHSAAGVGMAAIQIARHLGA- 1565
 QY 197 TKIVTVASK-----KHEKLLKS-----YGADVF-----DYHDGV- 227
 1566 -EVYATASDDKMGQAVELSRHEHLASSRTCDFOGQFLGATGGRVDVYVLSLAGFADASLR 1624
 DB 228 -----IE-----QIKSKYENLGHVIDAV-GSEDSIPAYVTVADSLPATL 267
 1625 MLPRGGRFLEKTVDRDVEVADAHGVSYAFDVEAGPORIGEMHLELVELEGRVL 1684
 QY 268 EYVPMT-----IESIEPEIR-----KONVKIDITL--LYRASGOEILLGAT 306
 DB 1685 EPLVTVAMDVROAPEALRHLSQLRHVGLVLTMPVMDAGIVLVGTG 1733

RESULT 13

US-09-370-700-3

Sequence 3, Application US/09370700

Patent No. 6274350

GENERAL INFORMATION:

APPLICANT: Baltz, Richard H.
 APPLICANT: Broughton, Mary C.
 APPLICANT: Crawford, Kathryn P.
 APPLICANT: Madduri, Krishnamurthy
 APPLICANT: Treadway, Patti J.
 APPLICANT: Turner, Jan R.
 APPLICANT: Waldron, Clive
 TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
 FILE REFERENCE: 50489 DIV1
 CURRENT APPLICATION NUMBER: US/09/370,700
 CURRENT FILING DATE: 1999-08-09
 EARLIER APPLICATION NUMBER: US 09/36987
 EARLIER FILING DATE: 1998-03-09
 NUMBER OF SEQ ID NOS: 39
 SOFTWARE: Patent in Ver. 2.0
 SEQ ID NO 3
 LENGTH: 2152
 TYPE: PRT
 ORGANISM: Saccharopolyspora spinosa
 US-09-370-700-3

Query Match 6.0%; Score 115.5; DB 4; Length 2152;

Best Local Similarity 22.6%; Pred. No. 0.039;
Matches 79; Conservative 48; Mismatches 129; Indels 93; Gaps 19;

QY 17 AVVKDVSVPELKEGTAIVKVEAVAGNPDMKHAIKIGEGSILGCDIAGTVKGPNA 76
DB 1419 ALVDEPTAALPDGDEVRIMRAGVAFRD-ALIALGMYPGVASISSEGAVVETGPGV 1477
QY 77 STDLKAGDTGPGFVHAGSQTDPKNGAFEAARYVPLPFYKSNLTHSTADEISEGPKNPE 136
DB 1478 -TGLAGDVRMGMI-----PK-----AFGL--AAVDHMTTRIPG--WSRA 1515
QY 137 SAASLPSLTAGVSLCHHLGSKMEHMPSTQHTHPLIWGATAVGQOLIQAQKINAY 196
DB 1516 RAASVIVELTAYVALVDLAGLR-----PG-----ESLVHSAAGVGMAIGARHLGA- 1565
QY 197 TKIVTVASK-----KHEKLLKS-----YQADVP-----DYHDAGV- 227
DB 1566 -EVYATASDDKQAVELSRHSLASSRTCDFEQOFLGATGGRGVVLSLAGEFPDASLR 1624
QY 228 -----IE-----QIKSKYPNLOH-VIDAV-GSEDSIPRAYKVTADSLPATLL 267
DB 1625 MLPGRGFLELKTVDYRDEVEVADAHFVSYQAFDIVEAGPQRTIGEMHELVLEFGRVL 1684
QY 268 EVVPMT---IESIPEIR-----KQNVKIDITL--LYRASGOEILLGAT 306
DB 1685 EPLPVTAAMDVRQAPFALRHLSQARHVQKLVLTWMPVWDAAGTVLVGTGT 1733

RESULT 14
US-08-804-227C-10
; Sequence 10, Application US/08804227C
; Patent No. 5876991

GENERAL INFORMATION:

APPLICANT: Dehoff, Bradley S.
APPLICANT: Kuhnstoss, Stuart A.
APPLICANT: Rostock, Paul R., Jr.
APPLICANT: Sutton, Kimberly L.
TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: THOMAS G. PLANT 1501
STREET: LILLY CORPORATE CENTER
CITY: INDIANAPOLIS
STATE: IN
COUNTRY: USA
ZIP: 46285

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: ASCII(DOS) Text only
CURRENT APPLICATION DATA:
FILING DATE: February 21, 1997
APPLICATION NUMBER: US/08/804,227C

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:

NAME: Plant, Thomas, G.
REGISTRATION NUMBER: 35,784
REFERENCE/DOCKET NUMBER: X-8231
TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-276-2459
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 3724 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-804-227C-10

Query Match 5.9%; Score 114; DB 2; Length 3724;
Best Local Similarity 23.7%; Pred. No. 0.13;
Matches 85; Conservative 48; Mismatches 139; Indels 86; Gaps 18;

QY 12 IEGDRAVVTDSVPELKEGTAIVKVEAVAGNPDMKHAIKIGEGSILGCDIAGTVK 71
DB 2955 LDGLALVPAEDAP-LEGGQVAVRAGVAFRD-ALIALGMYPGEAMGTGEGATVE-3012

QY 72 LCPNASTDLKAGDTGFGFVHAGSQTDPKNGAFEAARYVPLPFYKSNLTHSTADEISEGP 131
DB 3013 VGPVG- TGVAVDGDRVLGLWDG-----LGPL-----CVADHRLLP 3047

QY 132 VK---NFESASLPSLTAGVSLCHHLGSKMEHMPSTQHTHPLIWGATAVGQOLIQA 188
DB 3048 VPDGMSYQAASVPAVFLSAYYGLVTLAQLR-----PG-----ERYLVHAAAGVGMAAQAQ 3098

QY 189 VAKHINAYTKIVTASVSKHEKLLKSYGADVPDYHDAGVIEQIKSKYPLQVHTDVGSE 248
DB 3099 IARHLGA-EVLTATSPGKMDLIRAMGTD-----DHLASS 3132

QY 249 DSIPAEYKVT-ADSLPATLLEVPMTIESIPEIRKQNVKIDITL-LYRASGOEILLGAT 306
DB 3133 RTLDFATFTGADG--TSRADVV--LNSLTKEF-----VDASGLLRPGRFLELGT 3181

QY 307 REPASPEHEATVVKVYKFINPHLN-----GDHNNIKVRSNG-LDDVPALTGKIK 357
DB 3182 DV-RDPERIAABHPGVRYRAFPLNAGPDALGRLRLREIMDLFAAGVHPLPVVTHDVR 3238

RESULT 15
US-08-804-198-4
; Sequence 4, Application US/08804198
; Patent No. 5945320

GENERAL INFORMATION:

APPLICANT: Burgett, Stanley G.
APPLICANT: Kuhnstoss, Stuart A.
APPLICANT: Rao, Nagaraja R.
APPLICANT: Richardson, Mark A.
APPLICANT: Rostock, Paul R., Jr.
TITLE OF INVENTION: PLATENOILIDE SYNTHASE GENE
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: PAUL R. CANTRELL 1138
STREET: LILLY CORPORATE CENTER
CITY: INDIANAPOLIS
STATE: IN
COUNTRY: USA
ZIP: 46285

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: Macintosh
OPERATING SYSTEM: Macintosh 7.0
SOFTWARE: Microsoft Word 5.1
CURRENT APPLICATION DATA:
FILING DATE:
APPLICATION NUMBER: US/08/804,198

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:

NAME: Cantrell, Paul R.
REGISTRATION NUMBER: 36,470
REFERENCE/DOCKET NUMBER: P9113
TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-276-3885
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 3724 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-804-198-4

Query Match 5.9%; Score 114; DB 2; Length 3724;
Best Local Similarity 23.7%; Pred. No. 0.13;
Matches 85; Conservative 48; Mismatches 139; Indels 86; Gaps 18;

QY 12 IEGDRAVVTDSVPELKEGTAIVKVEAVAGNPDMKHAIKIGEGSILGCDIAGTVK 71
DB 2955 LDGLALVPAEDAP-LEGGQVAVRAGVAFRD-ALIALGMYPGEAMGTGEGATVE-3012

```
Db      2955 LDGLALVPAPDAAP-LEPGQVAVAAAGVNRD-ALLAGMTPEAEKMTGAGTVE 3012
QY      72  LGPNASTDLKVGDPFGFVHGASQTDPKNGAFAYARVPPLFYKSNLTHSTADEISEGP 131
Db      3013 VGPQV--TGVAVGDRLVGLMDG-----LGPL-----CVADHRLAP 3047
QY      132 VK--NFEASASLPVSLTTAGVSLCHHLGSKMHNHSTPQHTHPLIHWGATAVGQOLIQ 188
Db      3048 VPDGWSYAQAASVPAVFLSAYYGLVTLGLR---PG---BRVLVHAAAGGVGMAAVQ 3098
QY      189 VAKHINAYTKIVTVASKKHEKLSYGADVDFYHDAGVLEQIKSKYPNLQHVIDAVGSE 248
Db      3099 IAHHLGA--EVLATASGKMDALRAMGITD-----DHLASS 3132
QY      249 DSIPEAYKT-ADSLPATLLEVPMTIESIPEIRKDNKIDITL-LYRASGOEILLGAT 306
Db      3133 RTLDFAITAFGADG--TSRADV--LNSLTKEF-----VDASLGILLRPGGRFLELGKT 3181
QY      307 RFPASPEYHEATYKVFYFIPHLNN-----GDIHNNIKVPSNG-LDDVPALTEGIX 357
Db      3182 DV-RDPERIAAEHPGVRYRAFDLNEAGPDALGRLLREIMDLFAAGVLAHPLPVVTHDVR 3238
```

Search completed: June 17, 2003, 13:01:58
Job time : 21 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2003, 10:49:17 ; Search time 140 Seconds
(without alignments)
282.563 Million cell updates/sec

Title: US-10-081-644-2

Perfect score: 1920

Sequence: 1 MSVPTTKAVIIEGDKAVK.....ALTEGKEKKNKVKYVARL 370

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Reached: 408643 seqs, 106915682 residues

Number of hits satisfying chosen parameters: 408643

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA.*

- 1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1920	100.0	370	US-10-081-644-2	Sequence 2, Appli
2	1065.5	55.5	376	US-10-081-644-4	Sequence 4, Appli
3	1021.5	53.2	377	US-10-081-644-6	Sequence 6, Appli
4	1015.5	52.9	368	US-10-081-644-8	Sequence 8, Appli
5	266	13.9	318	US-10-137-036-66	Sequence 66, Appli
6	265	13.8	322	US-09-815-242-10703	Sequence 10703, A
7	206	11.7	338	US-09-815-242-10944	Sequence 10944, A
8	168	8.8	337	US-09-738-626-6825	Sequence 6825, A
9	155	8.1	317	US-10-137-036-65	Sequence 65, Appli
10	149	7.8	318	US-09-738-626-3730	Sequence 3730, Ap
11	144.5	7.5	352	US-09-908-744-69	Sequence 69, Appli
12	143.5	7.5	350	US-09-847-208-53	Sequence 53, Appli
13	139.5	7.3	337	US-09-815-242-11926	Sequence 11926, A
14	134.5	7.0	306	US-09-738-626-6782	Sequence 6782, Ap
15	131	6.8	333	US-09-815-242-5629	Sequence 5629, Ap
16	131	6.8	335	US-09-815-242-12251	Sequence 12251, A
17	131	6.8	377	US-10-160-501-2	Sequence 2, Appli
18	131	6.8	377	US-09-838-573-2	Sequence 2, Appli
19	130.5	6.8	340	US-09-971-361-10	Sequence 10, Appli

20	125.5	6.5	324	10	US-09-815-242-5117	Sequence 5117, Ap
21	122.5	6.4	336	1	US-08-976-063C-20	Sequence 20, Appli
22	122	6.4	346	10	US-09-815-242-10162	Sequence 10162, A
23	121	6.3	340	9	US-09-738-626-6285	Sequence 6285, Ap
24	119.5	6.2	352	9	US-09-908-744-52	Sequence 52, Appli
25	119	6.2	350	9	US-09-908-744-53	Sequence 53, Appli
26	114.5	6.0	1616	9	US-09-712-363-262	Sequence 262, App
27	112	5.8	264	10	US-09-815-242-11710	Sequence 11710, A
28	112	5.8	336	10	US-09-815-242-13832	Sequence 13832, A
29	111	5.8	346	10	US-09-815-242-13437	Sequence 13437, A
30	111	5.8	346	10	US-09-815-242-13656	Sequence 13656, A
31	111	5.7	350	9	US-09-908-744-51	Sequence 51, Appli
32	109.5	5.8	342	10	US-09-815-242-10958	Sequence 10958, A
33	107.5	5.6	356	10	US-09-925-300-1326	Sequence 1326, Ap
34	107.5	5.6	3815	9	US-09-808-880-3	Sequence 3, Appli
35	105.5	5.5	283	10	US-09-925-302-693	Sequence 693, App
36	105.5	5.5	1007	10	US-09-765-272-216	Sequence 216, App
37	104.5	5.4	2563	9	US-09-836-705-46	Sequence 46, Appli
38	103.5	5.4	369	10	US-09-854-122-40	Sequence 40, Appli
39	103.5	5.4	385	9	US-09-984-334-1	Sequence 1, Appli
40	103.5	5.4	455	10	US-09-775-009-7	Sequence 7, Appli
41	103	5.4	357	10	US-09-947-027-9	Sequence 9, Appli
42	103	5.4	357	12	US-10-091-009-9	Sequence 9, Appli
43	102.5	5.3	457	9	US-10-002-245-4	Sequence 4, Appli
44	102	5.3	302	10	US-09-815-242-11730	Sequence 11730, A
45	101	5.3	362	10	US-09-947-150-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1	US-10-081-644-2	Application US/10081644
Sequence 2, Appli	US-10-081-644-2	Publication No. US20020192782A1
GENERAL INFORMATION:		
APPLICANT:	Yamamoto, Hiroaki	
APPLICANT:	Kimoto, No. US20020192782A1, Hiro	
TITLE OF INVENTION:	NOVEL ENZYME REDUCTASES, METHODS FOR	
TITLE OF INVENTION:	PRODUCING SAME, AND METHODS FOR SELECTIVELY REDUCING A	
TITLE OF INVENTION:	CARBON-CARBON DOUBLE BOND OF AN ALPHA, BETA-UNSATURATED KETONE	
FILE REFERENCE:	06501-100001	
CURRENT APPLICATION NUMBER:	US/10/081,644	
CURRENT FILING DATE:	2002-02-21	
PRIOR APPLICATION NUMBER:	JP 2001-49363	
PRIOR FILING DATE:	2001-02-23	
NUMBER OF SEQ ID NOS:	29	
SOFTWARE:	FastSeq for Windows Version 4.0	
SEQ ID NO:	2	
LENGTH:	370	
TYPE:	PRT	
ORGANISM:	Kluyveromyces fragilis	
US-10-081-644-2		
Query Match	100.0%	Score 1920; DB 9; Length 370;
Best Local Similarity	100.0%	Pred. NO. 5.4e-160;
Matches	370;	Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	1	MSVPTTKAVIIEGDKAVKTDVSVPELKEGTALVKEVAVAGNPTDMKHIAKIGPEGSI 60
DB	1	MSVPTTKAVIIEGDKAVKTDVSVPELKEGTALVKEVAVAGNPTDMKHIAKIGPEGSI 60
QY	61	LCGDIAGTVVKIGPNASTDLYKGTGFGFVHCASGTDPRNGAFAYARVYPLPFYSNLT 120
DB	61	LCGDIAGTVVKIGPNASTDLYKGTGFGFVHCASGTDPRNGAFAYARVYPLPFYSNLT 120
QY	121	HSTADEISGPKVKNESASLPVSLTTAGVSLCHLIGSKMEWHPEPTPOHTPLILWGAT 180
DB	121	HSTADEISGPKVKNESASLPVSLTTAGVSLCHLIGSKMEWHPEPTPOHTPLILWGAT 180
QY	181	AVGGQILGVAKHINAYTKIVASKKHEKLLKSYGADVDFVHDAGVLEQISKYPNLOH 240
DB	181	AVGGQILGVAKHINAYTKIVASKKHEKLLKSYGADVDFVHDAGVLEQISKYPNLOH 240

Db 181 AVGOQOLIIVAKHINATKIIVTASKEKELKSYGADVDFYHDAGVIEQIKSKYPNLOH 240
 QY 241 VIDAVSESDSIPPAKYVTADSLPATLLEVPMTIESIPBEIRKDNVKIDITLLYRASGOE 300
 Db 241 VIDAVSESDSIPPAKYVTADSLPATLLEVPMTIESIPBEIRKDNVKIDITLLYRASGOE 300
 QY 301 ILIGATRPASPEYHEATVKEVFIPNHLNNGDIHHNNIKVPSNGLDVDPALTEGIEKEK 360
 Db 301 ILIGATRPASPEYHEATVKEVFIPNHLNNGDIHHNNIKVPSNGLDVDPALTEGIEKEK 360
 QY 361 NKXKXKVAVL 370
 Db 361 NKXKXKVAVL 370

RESULT 2

US-10-081-644-4
 ; Sequence 4, Application US/10081644
 ; Publication No. US20020192782A1

GENERAL INFORMATION:

APPLICANT: Yamamoto, Hiroaki

TITLE OF INVENTION: NOVEL ENONE REDUCTASES, METHODS FOR

TITLE OF INVENTION: PRODUCING SAME, AND METHODS FOR SELECTIVELY REDUCING A

TITLE OF INVENTION: CARBON-CARBON DOUBLE BOND OF AN ALPHA, BETA-UNSATURATED KETONE

FILE REFERENCE: 06501-100001

CURRENT FILING DATE: US/10/081,644

PRIOR FILING DATE: 2002-02-21

PRIOR FILING DATE: 2001-02-23

NUMBER OF SEQ ID NOS: 29

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 4

LENGTH: 376

TYPE: PRN

ORGANISM: Saccharomyces cerevisiae

US-10-081-644-4

Query Match
 Best Local Similarity 54.0%; Pred. No. 3.8e-85;
 Matches 201; Conservative 59; Mismatches 109; Indels 3; Gaps 2;

QY 2 SVPTTQKAVIIEGDAVAVKTVSVPELKEGTALVKEVAAGNPTDMKHIAVKGPEGSIL 61
 Db 4 SIBETKAVAVIENGKAVKODIPPELEEGFVLKTVAVAGNPTDMKHIDFKIGPQALL 63
 QY 62 GCGLAGTVVVLGPNA-STDLKVGDTGFGFVHGASQDTPKNGAFAYARVPLPYK--SN 118
 Db 64 GCDAAGQIVLGPVNDARPAIGYITGVHIGASVRFPSNGAFAYSAISSETAYKPAR 123
 QY 119 LTHSTADEISEGVPKNFESASLPSVSLTTAGVSLCHHLSGKMEHPSPTQHTPLLIW 178
 Db 124 FRICGKDXLPEGVYKSLGAVSLPSVSLTTAGMLTSHFGIDMTWKPKARDQPIILFW 183
 QY 179 AIVAGQOLIIVAKHINATKIIVTASKEKELKSYGADVDFYHDAGVIEQIKSKYPN 238
 Db 184 AIVAGQOLIIVAKHINATKIIVTASKEKELKSYGADVDFYHDAGVIEQIKSKYPN 243
 QY 239 QHVIDAVSESDSIPPAKYVTADSLPATLLEVPMTIESIPBEIRKDNVKIDITLLY 298
 Db 244 PYLDVCSNTEITQOVYKCADDDATVVDIVLTVLEEDRQVVSIEGTLVILIG 303
 QY 299 QEILLAGTRPASPPEYHEATVKEVFIPNHLNNGDIHHNNIKVPSNGLDVDPALTEG 358
 Db 304 NDVPFGFTLPPADPEYEAIAIKFIKIPNKHINDEIHIIIVKVKNGLDIPOLDIDIK 363
 QY 359 GKXKXKVAVL 370
 Db 364 GRNKGKLVAVL 375

RESULT 3

US-10-081-644-6
 ; Sequence 6, Application US/10081644
 ; Publication No. US20020192782A1

GENERAL INFORMATION:

APPLICANT: Yamamoto, Hiroaki

TITLE OF INVENTION: NOVEL ENONE REDUCTASES, METHODS FOR

TITLE OF INVENTION: PRODUCING SAME, AND METHODS FOR SELECTIVELY REDUCING A

TITLE OF INVENTION: CARBON-CARBON DOUBLE BOND OF AN ALPHA, BETA-UNSATURATED KETONE

FILE REFERENCE: 06501-100001

CURRENT FILING DATE: US/10/081,644

PRIOR FILING DATE: 2002-02-21

PRIOR FILING DATE: 2001-02-23

NUMBER OF SEQ ID NOS: 29

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 6

LENGTH: 377

TYPE: PRN

ORGANISM: Saccharomyces cerevisiae

US-10-081-644-6

Query Match
 Best Local Similarity 51.7%; Pred. No. 2.7e-81;
 Matches 193; Conservative 67; Mismatches 110; Indels 3; Gaps 2;

QY 1 MSVPTTQKAVIIEGDAVAVKTVSVPELKEGTALVKEVAAGNPTDMKHIAVKGPEGS 60
 Db 4 VAIPETKAVIIEGDAVAVKTVSVPELKEGTALVKEVAAGNPTDMKHIAVKGPEGS 63
 QY 61 LGCDIAGTVVVLGPNA-STDLKVGDTGFGFVHGASQDTPKNGAFAYARVPLPYK-- 117
 Db 64 LGCDIAGTVVVLGPNA-STDLKVGDTGFGFVHGASQDTPKNGAFAYARVPLPYK-- 123
 QY 118 LTHSTADEISEGVPKNFESASLPSVSLTTAGVSLCHHLSGKMEHPSPTQHTPLLI 177
 Db 124 ELKPEGEDVLPAGVPVSLGAVSLPSVSLTTAGVSLCHHLSGKMEHPSPTQHTPLLI 183
 QY 178 GATVAGQOLIIVAKHINATKIIVTASKEKELKSYGADVDFYHDAGVIEQIKSKYPN 237
 Db 184 GATVAGQOLIIVAKHINATKIIVTASKEKELKSYGADVDFYHDAGVIEQIKSKYPN 243
 QY 238 LGHVIDAVSESDSIPPAKYVTADSLPATLLEVPMTIESIPBEIRKDNVKIDITLLY 297
 Db 244 ISYLVDCVANODTLQOVYKCADDDATVVDIVLTVLEEDRQVVSIEGTLVILIG 303
 QY 298 GORILLAGTRPASPPEYHEATVKEVFIPNHLNNGDIHHNNIKVPSNGLDVDPALTEG 357
 Db 304 GHEVPPGNITLPPADPEYEAIAIKFIKIPNKHINDEIHIIIVKVKNGLDIPOLDIDIK 363
 QY 358 EGKXKXKVAVL 370
 Db 364 YGKXGKLVAVL 376

RESULT 4

US-10-081-644-8
 ; Sequence 8, Application US/10081644
 ; Publication No. US20020192782A1

GENERAL INFORMATION:

APPLICANT: Yamamoto, Hiroaki

TITLE OF INVENTION: NOVEL ENONE REDUCTASES, METHODS FOR

TITLE OF INVENTION: PRODUCING SAME, AND METHODS FOR SELECTIVELY REDUCING A

TITLE OF INVENTION: CARBON-CARBON DOUBLE BOND OF AN ALPHA, BETA-UNSATURATED KETONE

FILE REFERENCE: 06501-100001

CURRENT FILING DATE: US/10/081,644

PRIOR FILING DATE: 2002-02-21

PRIOR FILING DATE: 2001-02-23

NUMBER OF SEQ ID NOS: 29

SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 8
LENGTH: 368
TYPE: PRT
ORGANISM: Saccharomyces cerevisiae
US-10-081-644-8

Query Match 52.9%; Score 1015.5; DB 9; Length 368;
Best Local Similarity 53.0%; Pred. No. 8.7e-81;
Matches 194; Conservative 60; Mismatches 109; Indels 3; Gaps 2;

QY 8 KAVIIEGDAVAVKTDVSVPELKEGTALVKEVAAGNPTDMKIIAYKIGEGSILGCDIAG 67
DB 2 KAVIIEGDAVAVKTDVSVPELKEGTALVKEVAAGNPTDMKIIAYKIGEGSILGCDIAG 61
QY 68 TVVKGAGNNS-TDLKVGDTGFGFVHGASQTDPRKNGAFAYARVYPLPYKS--NLTHSTA 124
DB 62 QIVKLGPAVDPKDFSIGDYIYGFHGSVYRFPNGAFAYSAISTVAVYKSPNELKFLGE 121
DB 125 DEISGVPKNGFSAASLPVSLTTAGVSLCHHIGSKKEMHSPPOHTHPLILMGATAVGO 184
DB 122 DVLPAPVPSLBEAATIPVSLTTAGVSLYTNLGLNLMKESPSTPQNNGPILLMGATAVGO 181
QY 185 QLIQVAKHINAYTKIYTVASKKHEKLKSYGADVDVHDAGVIEQIKSKYPMLOHVIDA 244
DB 182 SLIQLANKNGFTKIIVASRKGKELKKEYGADQLFDYHDIDVVEQIKKKNNISLVLC 241
QY 245 VQSEDSIPRAYKVTADSLPATLLEVPMTIESIPESIRQNKYKIDITLLYRASGGEIILG 304
DB 242 VANQNTLQOVYKCAADKODATVVELTNTLEENVKENRRQNVITDRTRLYSIGGHEVPRG 301
QY 305 ARFPASPEYHEATVYFVFINPLANGDIHNNIIVFNGLDVVALTEGKESKKNV 364
DB 302 GTFPPADPEARRATEFVFNKISDGLIHIPARVKNGLYDVERILIEDIKIGNSGE 361
QY 365 KYVAVL 370
DB 362 KLVAVL 367

RESULT 5
US-10-137-036-66
Sequence 66, Application US/10137036
Publication No. US20030101478A1
GENERAL INFORMATION:
APPLICANT: Petera, Ranjan
APPLICANT: Rice, Stephen
APPLICANT: Easleyton, Claire
APPLICANT: Lasham, Annette
APPLICANT: Wood, Marion
APPLICANT: Wieser, Elizabeth
TITLE OF INVENTION: Compositions and Methods for the
FILE REFERENCE: 11000.1036C4
CURRENT APPLICATION NUMBER: US/10137,036
PRIOR FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: PCT/NZ 01/00115
PRIOR FILING DATE: 2001-06-20
PRIOR APPLICATION NUMBER: U.S. No. US20030101478A1 09/724,624
PRIOR FILING DATE: 2000-11-28
PRIOR APPLICATION NUMBER: U.S. No. US20030101478A1 09/598,401
PRIOR FILING DATE: 2000-06-20
PRIOR APPLICATION NUMBER: PCT/NZ00/00018
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: U.S. No. US20030101478A1 60/146,591
PRIOR FILING DATE: 1999-07-30
PRIOR APPLICATION NUMBER: U.S. No. US20030101478A1 09/276,559
PRIOR FILING DATE: 1999-03-25
NUMBER OF SEQ ID NOS: 143
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 66
LENGTH: 318
TYPE: PRT

ORGANISM: Eucalyptus grandis
US-10-137-036-66

Query Match 13.9%; Score 266; DB 9; Length 318;
Best Local Similarity 33.6%; Pred. No. 3.1e-15;
Matches 91; Conservative 31; Mismatches 93; Indels 56; Gaps 11;

QY 3 VTTOKAVIIL--EGGKAVV--KTDSVPELKEGTALVKEVAAGNPTDMKIIAYKIGPE 57
DB 7 IPTKKKAVYREHGVAVNLGLDPELKVPELQEGVAVKVLAAALNPIDTARVKVITLP 66
QY 58 G----SILGCDIAGTVKIGPNASTDLVKGDTFGFGVHGASQTDPRKNGAFAYARVYPL 113
DB 67 GSSLAVAGGYDLAVGVVAVKGEVK--ELKVGSEVYGFMFHAK---KDGTLAEYAAVESSF 121
QY 114 FKSMLTSTADEISEGVK--NESAASLPVSLTT-----AGVSLCHHIGSKKEMHPS 165
DB 122 -----LATLKKRKRFGESASLPVVIQRAYGGLERAGLS----- 154
QY 166 TPQHTHPLILMGATAVAGQOLIQVAKHINAYTKIYTVASKKHEKLKSYGADVDVHDYDA 225
DB 155 ---HKSLLVLAGAGCGVTLIIQLKKEVFGASRVATSTGKLELKSGLADLAIDYKV 211
QY 226 GVIEQIKSKYPMLOHVIDAVGSEDSIPRAYK 256
DB 212 N-FEDLPEKF--DVYDVTGEBERAAKAYK 238

RESULT 6
US-09-815-242-10703
Sequence 10703, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haseibeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA 011A
CURRENT APPLICATION NUMBER: US/09/815,242
PRIOR FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 10703
LENGTH: 322
TYPE: PRT
ORGANISM: Enterococcus faecalis
US-09-815-242-10703

Query Match 13.8%; Score 265; DB 10; Length 322;
Best Local Similarity 30.4%; Pred. No. 3.8e-15;
Matches 106; Conservative 42; Mismatches 125; Indels 76; Gaps 15;
QY 8 KAVIIE--GDKAVV--KTDSVPELKEGTALVKEVAAGNPTDMKIIAYKIGEG----- 58

```
Db      11 KAVINQYSGKEVLEAEVTLPELSEHQVKEVATSNPLDWLT-----REGYIKOMF 64
Qy      59 -----SILGCDIAGTVVVKLGPNASTDLKVGDTGFGFVHGASQTPDKNGAFAEYARVPL 113
Db      65 DMSRPILGMDVAGVITVEVSGV-TDWOVGDVYF-----ARPETTRGTAYEV---T 112
Qy      114 FYKSNLTHSTADEISEGPNKVPESASLPVSLTTAGVSLCHHLSKMEWHPSTPQHTHPL 173
Db      113 IVDHLLAKIPETIS-----FEBAAVPLAGLTAMQALPDHGLK-----EGETV 157
Qy      174 LINGATAVGQOLIOVAKHINAYTKIVTASKEHEKLSYGADVDYDAGVIEQI-- 231
Db      158 LHHAGAGGVGYALQILKEAGAH--VITTSARKHALLKIGADEVIDYHTTFAEVLTD 215
Qy      232 -----KSKY-----PNLOHVITAVGSEDSIPEAVK-VTADSLPATLLEVPMT 273
Db      216 VDLVFDITMGGEVQKNSDVLKPNTRGLVSVIGIEDKOLAEKVSASIS---WLOPNGEQ 272
Qy      274 IESIPETIRKONVKIDITLLYRASGQILGATRPSPPEYHEATVKEV 322
Db      273 LQKIADLMAAGKVSIIIGEVFPFSRQ-----GIYDAHALSETHHAVGKIV 317

RESULT 7
US-09-815-242-10944
; Sequence 10944, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815, 242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191, 078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206, 848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207, 727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242, 578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253, 625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257, 931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269, 308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 10944
; LENGTH: 338
; TYPE: PR1
; ORGANISM: Enterococcus faecalis
US-09-815-242-10944

Query Match      10.7%; Score 206; DB 10; Length 338;
Best Local Similarity 28.1%; Pred. No. 6,1e-10;
Matches 83; Conservative 33; Mismatches 87; Indels 86; Gaps 16;

Qy      7 QKAVIIEGDRAVYKTVDSVPELKEGTALVKEAVAGNPTDMKH-----IAKIGPEG 58
Db      12 QKELAIR-----EVPLPTIHDNDVLTAKIASINPDLTKTKDGKVKMLNYQM---P 60
Qy      59 SILGCDIAGTVVVKLGPNASTDLKVGDTGFGFVHGASQTPDKN--GAFAEYARV--KPPPL 114
```

```
Db      61 LILGSPFAGIIVSVGKKVO-NFRLGDVAVGRV-----PKRVRGFAEYAVINDQAAM 112
Qy      115 YKSNLTHSTADEISEGPNKVPESASLPVSLTTAGVSLCHHLSKMEWHPSTPQHTHPL 174
Db      113 KPNKLT-----FEBAATPL-----VGLTSQALHDMNVPG-----QKVL 149
Qy      175 IWCATAVGQOLIOVAKHINAYTKIVTASKEHEKLSYGADVDYDAGVIEQIKSK 234
Db      150 IQGSGGIGITIALQILAKLAGAY--VATTSKKNKEWQALGADVIDYRTON-FEEVLSD 206
Qy      235 YPNLOHVITAVGSEDSIPEAKVTADSLPATLLE-----VVP-----MTIESIPEE 280
Db      207 Y--DYVFDITMG-----GTLIEKAFSVKPOGKVVTLISGLPNE 241

RESULT 8
US-09-738-626-6825
; Sequence 6825, Application US/09738626
; Patent No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAMA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738, 626
; PRIOR FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 6825
; LENGTH: 337
; TYPE: PR1
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-6825

Query Match      8.8%; Score 168; DB 9; Length 337;
Best Local Similarity 29.4%; Pred. No. 1.3e-06;
Matches 69; Conservative 31; Mismatches 93; Indels 42; Gaps 11;

Qy      2 SVPTQKAVIIEGDRAVYKTVDSVPELKEGTALVKEAVAGNPTDMKHAYIGPE--GS 59
Db      20 STPSDQDFLID-----VEVDRPTPGPHDLVHIEAVSINPVDTK--VRMAGKQHXK 71
Qy      60 ILGCDIAGTVVVKLGPNASTDLKVGDTGFGFVHGASQTPDKNGAFAEYARVPLFYKSNL 119
Db      72 ILGPDAGEVAVVAGVGV-TLFNVGDK--VFVAGSNQ---RPSGNAEY----- 112
Qy      120 THSTADE--ISEGPNK-NFESASLPVSLTTAGVSLCHHLSKMEWHPSTPQHTHPL 176
Db      113 --QVDERLVGHAPQSLGANDAAALPVALTMBESLFDRLG-----VTOSTGTTLVL 163
Qy      177 GGATAVGQOLIOVAKHINAYTKIVTASKEHEKLSYGADVDYDAGVIEQI 231
Db      164 GSGGGVPSALIQILARLTGLKVAATASRPESQEWYTKLGAHEVID-HSKDLSBOI 217

RESULT 9
US-10-137-036-65
; Sequence 65, Application US/10137036
```

```
Publication No. US20030101478A1
GENERAL INFORMATION:
APPLICANT: Perera, Ranjan
APPLICANT: Rice, Stephen
APPLICANT: Bagleton, Claire
APPLICANT: Lasham, Annette
APPLICANT: Wood, Marion
APPLICANT: Visser, Elizabeth
TITLE OF INVENTION: Compositions and Methods for the
FILE REFERENCE: 11000.1036C4
CURRENT APPLICATION NUMBER: US/10/137,036
PRIOR FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: PCT/NZ 01/00115
PRIOR FILING DATE: 2001-06-20
PRIOR APPLICATION NUMBER: U.S. No. US20030101478A1 09/724,624
PRIOR FILING DATE: 2000-11-28
PRIOR APPLICATION NUMBER: U.S. No. US20030101478A1 09/598,401
PRIOR FILING DATE: 2000-06-20
PRIOR APPLICATION NUMBER: PCT/NZ00/00018
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: U.S. No. US20030101478A1 60/146,591
PRIOR FILING DATE: 1999-07-30
PRIOR APPLICATION NUMBER: U.S. No. US20030101478A1 09/276,599
PRIOR FILING DATE: 1999-03-25
NUMBER OF SEQ ID NOS: 143
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 65
LENGTH: 117
TYPE: PRT
ORGANISM: Eucalyptus grandis
US-10-137-036-65
```

```
Query Match      8.1%; Score 155; DB 9; Length 117;
Best Local Similarity 37.9%; Pred. No. 3,7e-06;
Matches 44; Conservative 16; Mismatches 42; Indels 14; Gaps 4;

QY      3 VPTTKAVIIE-----GDAVVKTDSVPELKEGTALVVEAVAGNPTDMKIHAYKIGPE 57
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      7 ITTKKAWYRRHGNVADVGLDPELAKVPELQEGVLYVLAALNPVDAAMKKGVTKLP 66
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY      58 G-----SILGCDIAGTVVVKLGPNASTDLKVGDTGFGFVHGSQTPDKNGAFAYARY 109
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      67 GFSLPVAVPYDLAGVYVVKVGREVK-ELKIGDEVYGMFPAK-----KDGTLAYAAV 117
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
RESULT 10
Sequence 3730, Application US/09738626
Publication No. US20020197605A1
GENERAL INFORMATION:
APPLICANT: NAKAGAMA, SATOSHI
APPLICANT: MIZOGUCHI, HIROSHI
APPLICANT: ANDO, SEIKO
APPLICANT: HAYASHI, MIKIRO
APPLICANT: OCHIAI, KEIKO
APPLICANT: YOKOI, HARUHIKO
APPLICANT: TATEISHI, NAKO
APPLICANT: SENOH, AKIHIRO
APPLICANT: IKEDA, MASATO
APPLICANT: OZAKI, AKIO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/09/738,626
CURRENT FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: JP 99/377484
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: JP 00/159162
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: JP 00/280988
PRIOR FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 7059
SOFTWARE: PatentIn ver. 3.0
```

```
SEQ ID NO 3730
LENGTH: 318
TYPE: PRT
ORGANISM: Corynebacterium glutamicum
US-09-738-626-3730
```

```
Query Match      7.8%; Score 149; DB 9; Length 318;
Best Local Similarity 27.0%; Pred. No. 5,4e-05;
Matches 67; Conservative 33; Mismatches 96; Indels 52; Gaps 11;
```

```
QY      16 KAVVKT-----DVSPELKEGTALVVEAVAGNPTDMKIIA--YKIGPGS--I 60
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      2 QAVTTEKRYTASLEIQLQEVVPVPTLKPGEVLVEVKAGVNRADLLQTOGNYPV-PAGASEI 60
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY      61 LGCODIAGTVVVKLGPNASTDLKVGDTGFGFVHGSQTPDKNGAFAYARYPPLFKSNLT 120
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      61 LGECAGVIVADGTGTG---VGDEVACLTG-----GGAYAVAV-----P 99
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY      121 HSTADEISEGVPKNFESASLPSVSLTTAGVSLCHHLGSKMEWHPEPTQHTHPLIWGAT 180
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      100 EGQLMPIEG--YSTVEASLIVEVACTVWSNIGMLAGLQKE-----DTFLIHGAG 148
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY      181 AVGOQLIGVANHINAYTKIVTASRKHEKLKSYGADVPFDYHAGVIBQIKSYPNLQH 240
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      149 GIGTFAIQWKGALGV-TVAVTAGSTKLTCKNLGADILIKYKEDFAVLKKNK---ADV 204
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY      241 VIDAVGSE 248
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      205 ILDIIGAK 212
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
RESULT 11
US-09-908-744-69
Sequence 69, Application US/09908744
Publication No. US20030068791A1
GENERAL INFORMATION:
APPLICANT: MASNIKOV Andrei
APPLICANT: Ojamo Heikki
APPLICANT: Povelainen Mira
APPLICANT: Gros Hakan
APPLICANT: Toivari Mervi
APPLICANT: Richard Peter
APPLICANT: Ruohonen Laura
APPLICANT: Koivuranta Kari
APPLICANT: Londeborough John
APPLICANT: Aristidou Aristos
APPLICANT: Penttil, Merja
APPLICANT: Plazenet-Menue Claire
APPLICANT: Deutscher Josef
TITLE OF INVENTION: Manufacture of Five-Carbon Sugars and Sugar Alcohols
FILE REFERENCE: 1427.0010005
CURRENT APPLICATION NUMBER: US/09/908,744
CURRENT FILING DATE: 2001-07-20
PRIOR APPLICATION NUMBER: PCT/FI01/00051
PRIOR FILING DATE: 2001-01-22
PRIOR APPLICATION NUMBER: US 09/488,581
PRIOR FILING DATE: 2000-01-21
PRIOR APPLICATION NUMBER: US 08/790,585
PRIOR FILING DATE: 1997-01-29
PRIOR APPLICATION NUMBER: US 08/368,395
PRIOR FILING DATE: 1995-01-03
PRIOR APPLICATION NUMBER: US 08/110,672
PRIOR FILING DATE: 1993-06-24
PRIOR APPLICATION NUMBER: US 07/973,325
PRIOR FILING DATE: 1992-11-05
NUMBER OF SEQ ID NOS: 70
SOFTWARE: PatentIn Ver. 3.0
SEQ ID NO 69
LENGTH: 352
TYPE: PRT
ORGANISM: Bacillus halodurans
US-09-908-744-69
```

[illegible]

```

RESULT 12
US-09-847-208-53
Sequence 53, Application US/09847208
Publication No. US20030082190A1
GENERAL INFORMATION:
APPLICANT: Saxon, Andrew
APPLICANT: Zhang, Ke
APPLICANT: Zhu, Daocheng
TITLE OF INVENTION: FUSION MOLECULES AND TREATMENT OF
FILE REFERENCE: US67,002A
CURRENT APPLICATION NUMBER: US/09/847,208
NUMBER OF SEQ ID NOS: 177
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 53
LENGTH: 350
TYPE: PR1
ORGANISM: Candida albicans (Yeast)
US-09-847-208-53

Query Match 7.5%; Score 143.5; DB 9; Length 350;
Best Local Similarity 22.4%; Pred. No. 0.00019;
Matches 94; Conservative 57; Mismatches 123; Indels 145; Gaps 24

Qy 3 VPTOKAVIIE--GPKAVKTDVSPBELKEGTAIVK-----AVAGNPDMWKNIAVK 53
Db 5 IPKQKAVVPDTNGQLVYK-DYRPTPKRPNLLHVKSSGCHTDHARKGDWP-LARK 62
Qy 54 I---GREGSLIGCDIAGTVKLGPNASTDLKVGD-TGEFVHGASQT-----DPK- 99
Db 63 LPLVGHGEG---AGVVVGMEGVNK-GWKIGDFAGIKWLNIGSCMSCGCGAEPK 114
Qy 100 -----NGAFAYAVRVVPLPYKSNLTHSTADEISEGVPKNFESAASLPVSLTTAG 149
Db 115 GEADLSGYTHDSFEQYA-----TDAY-----QAKIPAGTDLAN 150
Qy 150 VS--LCHNLG-----SKMEWNPSTPQHTHEPLLWGATAVGQOLIQVAKHINAYT 197
Db 151 VAPILCAVTVYKALKTADTLAAGCW-----VAISGAGGIGSLAVYARAMGLRV 200
Qy 198 KIVTVASKKHEGLTISYGADVDVFDHADGAVIEQIKSKVPNLDHVIDAVGSEDSIPEAYKV 257
Db 201 VAIDGDEKGE-FVASLGHNAVDF-----TKDKDIVEAYVK 236

```

QY 258 TATSLPATLTLEVPMTIESIPREIKRONVKXIDITLLY-RASGOSIILGAREFPA-----S 311

Db 237 ATDGGHGAINV-----SVSEK-----ALDQSEVEYRPRGKVLVG-----LPNNAKVTA 281

QY 312 PEHNATLVEFVFNIPHLNMGDIHNNIKVNSN-----GLDVPALATBGIKRGK 360

Db 282 PVF-DLVVVSIEIKSGYGNRKDTAEALDFSRGILKPRICIVGLSDLPFYFKLMEBGK 339

```

RESULT 13
US-09-815-242-11926
Sequence 11926, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselebeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
TITLE OF INVENTION: Prokaryotes
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815.242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 11926
LENGTH: 337
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-815-242-11926

```

```

Query Match      7.3%; Score 139.5; DB 10; Length 337;
Best Local Similarity 26.9%; Pred. No. 0.0004;
Matches    59; Conservative   35; Mismatches   90; Indels   35; Gaps   10;

QY      22 DVSVELKEG--TALVKTAEAVGNPTWKHIAKYIGE---GSILGCDIAGTVKLGPRVA 76
D8      20 DIELPEPTGPEDLLEVEFRAISVPVTK--VRQAQAEAGAAKVLDMAAGVRAVASEV 78
QY      77 STDLLKAGDTGFGEFGVGAQTDPKNKGAAEAYARVYPPLFYKSNLTHSTADEISEGVK-- 133
D8      79 SL-FRPGDRVM---YAADITRP--GSSNSELRHV-----DERIAGHLPSKL 117
QY      134 NESAASLPVSLTTGTGVSLSCHHLSKKMEWHSPQHHTPLLIMGATAVAGQOLIOVAKHI 193
D8      118 DPQAQAAPLELTITA---WELLERIQIDAEGRKDQSGLLVVGAAAGCVSIIIVLAROL 173
QY      194 NAYTKIVTVAASKHEKLKSYGADVDFPYHDAGVTEIQK 232
D8      174 TGINVIGTSRAETQAMVRDLGAHHVID-HGKIPLAEEIK 211

RESULT 14
US-09-738-626-6782
```

```

; Sequence 6782, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: YATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/37484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 6782
; LENGTH: 306
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
; US-09-738-626-6782

```

```

Query Match      7.0%; Score 134.5; DB 9; Length 306;
Best Local Similarity 25.4%; Pred. No. 0.0095;
Matches 82; Conservative 42; Mismatches 122; Indels 77; Gaps 18;

```

```

QY 16 KAVKTKDVPELKE-----GTALVKEAVAGNPTWKHIAVYKIGPEG----- 58
DB 3 KVVNSVEYGGPEMDELITNTPPQPGELGVKHAAGVNPIDMK---VNSGVAQTREL 59
QY 59 SLICDIAGTVKLGPNASTDLKVGDTGTFGVHGSQTPKNGAFEAARVYRPLFYKSN 118
DB 60 APGEASGIVTAVGQGVG-GEFVAGDPVLGLV-----APGVGVABDTLL----- 103
QY 119 LTHSTA---DETSEGVNKPESASLPSVLTGTAGVSLCHLHGS-KHEWHPSTPQHNPRL 174
DB 104 VAEISTVLKPELS-----FTDPAALPV---AGASA--YAGTHVLEBEG-----QSL 146
QY 175 IMGATVAVGQQLIQVAKHINAYTKITVASKHEKLLKSYGADVDVDYHDAGVIEQISK 234
DB 147 INAGGAGVGLMAQIGR-VHKF-QVVGVDHEDKRELIESGCA--IFVATGDVAABQVRL 202
QY 235 YPN-LOHVIDAVGSE-----DSIPAYVTVDSLPATLLEVPMTIESIPPE 280
DB 203 LPDGVVVPDLVGGELRVVAPLAKNPAPHVISAADAATVGLGQVLTARTPEWVGITGV 262
QY 281 IRKDNV--KIDITLLVRSQGEI 301
DB 263 VOIGLVDPKVDVTYPLEQAKAL 285

```

```

RESULT 15
US-09-815-242-5629
; Sequence 5629, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard

```

```

; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5629
; LENGTH: 333
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
; US-09-815-242-5629

```

```

Query Match      6.8%; Score 131; DB 10; Length 333;
Best Local Similarity 25.2%; Pred. No. 0.0022;
Matches 74; Conservative 38; Mismatches 126; Indels 56; Gaps 13;

```

```

QY 26 PELKEGTALVKEAVAGNPTWKHIAVYKIGPEGSLICDIAGTVKLGPNASTDLKVGDT 85
DB 26 PTPENDLIVKNSISVNPVDTKQROMEVTVPRVLGFDAGTVAIGPDV-TLSPGDV 84
QY 86 GGFVHGSQTPKNGAFEAARVYRPLFYKSNLTHSTADE--ISEGPVK-NFESAASLP 142
DB 85 VP-----YAG--SPNRQSNATYQLVSEAIKAPHNISANEAVSLP 124
QY 143 VSLTAVGSLCHLHGSKHEWHPSTPQHNPRLIMGATVAVGQQLIQVAKHINAYTKITV 202
DB 125 LGITAVETFPDTF--KISHNPS-ENIGKSVLIINGAGVGSIATQIAKRY-GLTVITTA 180
QY 203 ASKHEKLLKSYGADVDVDYHDAGVIEQISK-----YPN-----L 238
DB 181 SKQETTEMCEKMGADIVLN-HREDVLQFKEKEIPLVDYITCTVNTDLYNMIELIKPL 239
QY 239 QHVIDAVG-SEDSIPEAYVTVDSLPATLLEVPMTIESIPPEIRKDNVKIDIT 291
DB 240 GHTITIVAFNEQDNLKLSITFTHEFMFARP--IHRTPMIKQHEVLEDT 291

```

```

Search completed: June 17, 2003, 13:08:00
Job time : 140 secs

```

THIS PAGE BLANK (USPTO)

[illegible]

RESULT 2

A:Chemical protein YLR460C - Yeast (*Saccharomyces cerevisiae*)
 A:Alternate names: hypothetical protein I912.7
 C:Species: *Saccharomyces cerevisiae*
 C:Date: 30-Nov-1995 #sequence_revision 23-Feb-1996 #text_change 19-Apr-2002
 C:Accession: S59418
 R:Kirsten, U.
 submitted to the EMBL Data Library, March 1995
 A:Description: The sequence of S. cerevisiae cosmid 9122.
 A:Reference number: S59414
 A:Accession: S59418
 A:Molecule type: DNA
 A:Residues: 1-376 <KIR>
 A:Cross-references: EMBL:U2283; NID:92264349; PID:g717077; GSPDB:GN000000
 A:Experimental source: strain S288C (AB972)
 C:Genetics:
 A:Gene: MIPS:YLR460C
 A:Cross-references: SGD:S0004452
 A:Map position: 12R

Query Match

Matches 193; Conservative 67; Mismatches 110; Indels 3; Gaps 2;

1 MSVPPTQKAVIIEGDKAVKTDVSVPELKEGTALVKVEAVAGNPTDWKHIAYKIGPEGSI 60

DB 3 VALPEIMKAVIEDGKAVKEGIPIPELEEGFVLIKITLAVAGNPTDMAHIDYKIGPGGSI 62

61 LGCDIAGIVKLGPNAS-IDLKVGDIIGFVHGASQIDPKNGAFAEYARVYPLFYKS-- 117

03 D0CDA30Q1 VNLGFAVNFJWFSIGDI IIGF IHGSS VKF PSNGAFAEISALSI VVAHNSFN 12Z

[illegible][illegible]

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

1

[illegible]

2008 ACCEPTED MANUSCRIPT

303	СНЕВБЕСНИТЪ РАДЕВАКА ИТЕКЕТНПКІНДСОВЪТЪ ВЪИКАНОСІ СЪВЪНТИ КДЪТЪ	363
-----	---	-----

Ов: 358 EGKNKVKVABT. 370

Dh 363 YGKNSGEKLVAVI, 375

RESULT 3

S19414
 hypothetical protein YCR102c - yeast (*Saccharomyces cerevisiae*)
 C:Species: *Saccharomyces cerevisiae*
 C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 19-Apr-2002
 C:Accession: S19414
 R:Grivell, L.A.; de Haan, M.; Maat, M.J.
 submitted to the Protein Sequence Database, March 1992
 A:Reference number: S19412
 A:Accession: S19414
 A:Molecule type: DNA
 A:Residues: 1-368 <GRI>
 A:Cross-references: EMBL:X59720; NID:g1907116; PID:e264727; PID:g1907240; GSPDB:GN00003;
 C:Genetics:
 A:Gene: MIPS:YCR102c
 A:Cross-references: SGD:S0000699
 A:Map position: 3R

Best L

Oy 8 KAVIIEGDKAVVKTIDVSPELKEGTALVKVEAAVGNPTDWKHIAYKIGPESILGCDIAG 67
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 2 KAVIIEGDKAVVGEVPIPELEGFVLTKTLAANSTPDNAHIDYKKGPQSGILGCDAAG 61
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Oy 68 TVVVKLGPNAS--TDLKVGDTGFPGFYHGASQTDPKNKAFAEYARVPPLPYGS--NLHTSTA 124
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 62 QIVYDGAVDWPKPFISIDYIYGFTHGSSVFAPNSGAFAEYSALSTVAAYKSPNELKEJGE 121
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Oy 125 DEISEGGKVPFESASAALPVSLTTAGVSLCHHLGSKMHEWSTPOHTPHLLIWGATATVQG 184
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 122 DVPAGPVRSLSEGATIPVSLTTAGLVLTYNGLINKMEPSTPOGRNPFILLMGATAVQG 181
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Oy 185 QILOVAHNIAHYKTVVVASKHKEXLLKSYGADDPVRYHAGYEIOIKSYRYPLOAHIDA 244
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 182 SLTOLAKMLNGFTKIIVASRKREKLKEYGAOLFEYHDIIVAEQIKKHYYNNISTVDC 241
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Oy 245 VGSEDSIPEAYKYTAOSLPRTLLEVVPMTIESIPEELRKDNVADITLLVLRASGOETLLG 304

Db 242

QY 305 ATRFPASPEYHEATVKVKEINPHLNGDIHNMNIKVFNSGLDDVPALTEGIKEGKNKNV 364

Db 302 GITFPADPEARATEFVKFINPKISDGIHHIPARVYKNGLYDVPRILEDIKIGKNSGE 361

QY 365 KYVART 370

DB 362 KLVAVL 367

RESULT 4

oxidoreductase homolog [imported] - *Listeria monocytogenes* (strain EGD

```
C:\Date: 27-Nov-2001 #sequence revision 27-Nov-2001 #text change 14-Dec-2001
```

R; Glaser, P.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.

D.; Jones, L.M.; Karst, U.

A; Authors: Krefl, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maltourn

A;Title: Comparative genomics of *Listeria* species.

A;Accession: AE1151

A: Molecule type: DNA
B: Residues: 1-313 <GTA>

A:Experimental] source: strain EGD-e
A;CROSS-references: GB:NC_003210; PIDN:CAC98691.1; PID:g1641000z; GSFDDB:GNV01

C; Superf.

C;Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology

A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-343 <NOF>
A:Cross-references: GB:AE000736; NID:g2983763; PIDN:AAC07327.1; PID:g2983766; GB:AE00065
A:Experimental source: strain VFS
C:Genetics:
A:Gene: adh1
C:Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology
C:Keywords: NAD; oxidoreductase
F:26-332/Domain: long-chain alcohol dehydrogenase homology <LADH>

Query Match 10.1%; Score 193; DB 1; Length 343;
Best Local Similarity 28.7%; Pred. No. 6.8e-07;
Matches 77; Conservative 43; Mismatches 94; Indels 54; Gaps 14;

QY 8 KAVIIEG---DKAVKTVSVPELKEGRLVKEVAAGNPDP-WGHI-AYKIGP-SSI 60
DB 2 KAVIIEGFGIIEKLVKEDEFPKPEKDEVLIRKAVLNHLDIWMGALPYKPELPHI 61
61 ICGDIAGTVKLGPNASTDLKVG-----DTGFGFVHAGASQTD-----PKNG 101
DB 62 IGSIVSGVVEKVG-SLVKNVKEGEVITAPGLSCGVWCOCGRDNHCDYDLGLKNG 120
QY 102 APAEYARVYPLPYKSNLTHSTADEISEGPVKN--FESAASLPVSLTTAGVSLCHHLGSK 159
DB 121 GYAEYAVV-----PARNVIKKP-KNLSFEBAASYPLTEL-----VWNLVVK 162
QY 160 MEMHSTPQHTNHLIWGATAVGQOLIOVAKHINAYTKIIVYASKHEKLSYGADV 219
DB 163 AQIKRPS-----RVFIWASSGSGVAGIDLAKAFNAFV-ITTAGNEKAKCKEIGADLV 216
QY 220 FDYHDAGVIEQISKY-PNLOHYIDAVG 246
DB 217 FHHYKDVYKRAVEFKEGDVVVDHIG 244

RESULT 11

guinone oxidoreductase BH0935 [imported] - Bacillus halodurans (strain C-125)
C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C:Accession: G83766
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A>Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A:Reference number: A83650; M01D:20512582; PMID:11058132
C:Accession: G83766
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-332 <STO>
A:Cross-references: GB:AP001510; GB:BA000004; NID:g10173440; PIDN:BAB04654.1; GSPDB:GN00
A:Experimental source: strain C-125
C:Genetics:
A:Gene: BH0935
C:Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology

Query Match 9.8%; Score 187.5; DB 2; Length 322;
Best Local Similarity 27.0%; Pred. No. 1.6e-06;
Matches 79; Conservative 44; Mismatches 83; Indels 87; Gaps 14;

QY 8 KAVIIEG---GDKAVVKT-DVSVPELKEGRLVKEVAAGNPDP-----WGHIAY 52
DB 2 KAVVYQYGDPSVLKMDVPKPTIGPTDVLINVKAGSTNPVDYPRKGIQVDEGFPHIH 61
QY 53 KIGEGSILGCDIAGTVKLGPNASTDLKVGDTGFGFVHAGASQTDPRKNGAFAYARVYRP 112
DB 62 -----FDVAGEIVEIGSDLE-NWQVGDVMA-----TNIK-GASAEYALIPHH 102
QY 113 LFYKSNLTHSTABEISGPVKNFESAASLPVSLTTAGVSLCHHLGSKMEMHSTPQHTP 172
DB 103 LRPV--LPKSV-----SYEEGAALAMTWTWTLHLS-----FDRGLQKGET 141
QY 173 LIIWGATAVGQOLIOVAKHINAYTKIIVYASKHEKLSYGADVFDYHDAGVIEQIK 232

DB 142 VVIYGGSGAVGAAIQAKRAGA-TVITTAGREGKEIQAQAGADQVIFYKE----- 192
QY 233 SKYPMLOHYIDAVGSEDSIPEAYKXTADSLPTLLEVPMTIE-SIPEIRKD 284
DB 193 -----ESVVDVAVK-----ATNGQVPLIDMSLSENMAOD 223

RESULT 12

zinc-binding oxidoreductase Atu5447 [imported] - Agrobacterium tumefaciens (strain C58,
A:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 01-Feb-2002
C:Accession: AH3214
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.
erage, G.; Gillet, W.; Grant, C.; Genthner, D.; Kuyavian, T.; Levy, R.; Li, M.; McClell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A>Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; PMID:11743193
A:Accession: AH3214
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-334 <KUR>
A:Cross-references: GB:AE008687; PIDN:AAL46134.1; PID:g17743901; GSPDB:GN00188
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: Atu5447
A:Genome: plasmid
C:Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology

Query Match 9.7%; Score 187; DB 2; Length 334;
Best Local Similarity 27.5%; Pred. No. 1.8e-06;
Matches 90; Conservative 40; Mismatches 123; Indels 74; Gaps 14;

QY 8 KAVIIEGDAVVK-----TDVSPPELKEGRLVKEVAAGNPDPWK-----HIAVKI 54
DB 2 KAFVY--DKYKKKGRLADMPNPIVIGANDVLRVHATNLNLDKVDGERKLFPLPY- 58
QY 55 GEGSILGCDIAGTVKLGPNASTDLKVGDTGFGFVHAGASQTDPRKNGAEYARVYPLPY 114
DB 59 --PFLIHGDLAGTVIRVAGANV-RQFKGDEVF-----ARPDHVGTFABEIAV----- 105
QY 115 YKSNLTHSTADISEGPVKNFESAASLPVSLTTAGVSLCH-----HGSMEHNPSTPQHT 170
DB 106 -----DAADIALKPTSLSMEOAASIPVGLTAWQALIEVGKVGSK----- 147
QY 171 HPLIWGATAVGQOLIOVAKHINAYTKIIVYASKHEKLSYGADVFDYHDAGVIEQ 230
DB 148 --VFIOAGSGVGTPIQLAKHLGA--TVATTSANAMELVSLGADVVIDYKTOD-FEQ 202
QY 231 IKSXPYNLOHYIDAVGSEDSI-----PEAYKXTADSLPTLLEVPMTIE 275
DB 203 VLSGYDVLVNSQDADATLEKSLNVLPRGKLISISEPDAFARSLKINPLRFVVRMSR 262
QY 276 SIPEIRKDNVAKIDITLLY-RASGOEI 301
DB 263 GVLRKASRRGV--DYSFLEFMRAGQOL 287

RESULT 13

B90457
alcohol dehydrogenase (Zn containing) (adh-12) [imported] - Sulfolobus solfataricus
C:Species: Sulfolobus solfataricus
C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 03-Aug-2001
C:Accession: B90457
R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Aweez, M.J.; Chan-
dong, I.; Ueffing, A.C.; Kozera, C.W.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, P.
aret, R.A.; Regan, M.A.; Senses, C.V.; Van der Oost, J.
submitted to Genbank, April 2001
A:Description: Sulfolobus solfataricus complete genome.

A;Reference number: A99139
A;Accession: B90457
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-331 <KIR>
A;Cross-references: GB:AE006641; NID:g13816149; PIDN:AAK42913.1; GSPDB:GN00155
C;Genetics:
C;Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology

Query Match
Best Local Similarity 25.5%; Pred. No. 1.2e-05;
Matches 80; Conservative 51; Mismatches 107; Indels 76; Gaps 16;

QY 21 TDVSVPELKEGTALVKEAVAGNPTDKHI-AVKIGEGSILGCDIAGTVKLGPNASTD 79
DB 18 TDYKNPEIGNHBEVLIRKLAGVNVDYTYVERLKVNPILPHIPGVFSGEVKVGADHYS- 76
QY 80 LKVGAD--TGFGFV-----HGASQTPKKGAFKFAKRYVPELKYKSNLT 120
DB 77 VSVGDRTYITIGRIFDGTCDCMAGYETVCRNGRIGVANGVMAEYVAVEKVFYKLPNE 136
QY 121 HSTADEISEGVKNFESASLPVSLTTAGVSLCHLGSKMEHSPSTPOHTHPIILMGAT 180
DB 137 YT-----WEMGSSILTVAAALTA-----YHALKEAQLSPS-----QTLVVFQASG 174
QY 181 AVGGQOLIQVAKHINAYTKIYTVASKKEKLKSYGADVDVPHDAGVIEQIK-----SKYP 236
DB 175 NTGFMFLVQLGKKFGA--KVIAVSRK---SWLRGAGFVVDVNE--VEEKVKETINGKMA 227
QY 237 NLGHVIDAVGSE--DSIPEAKVTADSLP-ATLLEVPMTIESIPEIRKDNKIDITLL 293
DB 228 DV--VINSLSBQFWDKSPSVLGVKGLVTEGTL-----GANVVDLSOL 270
QY 294 YRASGCEILGATR 307
DB 271 Y--SKHISILGVNR 282

RESULT 14
T18230
alcohol dehydrogenase (EC 1.1.1.1) II - Yeast (Candida albicans)
C;Species: Candida albicans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 08-Dec-2000
C;Accession: T18230
R;Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, November 1998
A;Reference number: Z18831
Accession: T18230
Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-348 <BAR>
A;Cross-references: EMBL:AL033501; PIDN:CAA21988.1
C;Genetics:
A;Note: adh2
C;Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology
C;Keywords: alcohol metabolism; NAD; oxidoreductase; zinc
F;44,67,154/Binding site: zinc, catalytic (Cys, His, Cys) #status predicted

Query Match
Best Local Similarity 9.1%; Score 175.5; DB 2; Length 348;
Matches 98; Conservative 49; Mismatches 138; Indels 113; Gaps 19;

QY 1 MSVPTTQKAVIIE--GDKAVVKTDSVPELKEGTALVKEAVAGNPTD--WK---HIAY 52
DB 1 MSVPTTQKAVIFENNGKLEIK-DIPVPRKAMLLINVKYSGVCHTDLAKMGDWLAL 59
QY 53 KI-----GPEGSILGCDIAGTVKLGPNASTDLKVG--TGFGFVHGA-----SQTDPK 99
DB 60 KLPLVGHGEG-----AGVVVALGENVK--GMKVGDAYGVMLNGSLCINCEYCSGGAEPN 111
QY 100 -----NGAFAYAVVPPPLFYKSNLHSTADEISEGVKNFESASLPVSLTTA 148

DB 112 CAEADLSGTTIHDSGFQYA-----TDAY-----QAAKIPACTDLA 147
QY 149 GVS--LCHNLGSKMEHNPSTPOHTHPIILMGATAVGQOLIQVAKINAYTKIYVASKK 206
DB 148 NVAPILCAGVTYVYKALKTEALEAGQWVAISGAAGGSLAVQYAKM--GTVLALIDGSD 206
QY 207 HEKLLKSYGADVFDY--HDAGVIEQIKSKYPLNQ--VIDAVGSEDSIPEAKYKTADSLA 264
DB 207 KGEFVKSIAEFTPIDTKEDVVEAVKAKATNGGPHGVIVSVSRALIGQS----- 256
QY 265 TLEVPMTIESIPEIRKDNKIDITLLYRASGCEILGATRFP--SPEYHEATVK 320
DB 257 -----TEYVRLIGKVLVG---LPGAALSTPVPFAVVK 287
QY 321 FYKFINPHLNGDIHNNIKVPSNGDDVPALTEGIKE 358
DB 288 TIQKGSYVGNKRDTEAVDFFRGLIKPKIVGLSE 325

RESULT 15
AG3182
zinc-binding dehydrogenase Atu5188 [imported] - Agrobacterium tumefaciens (strain C58, D
C;Species: Agrobacterium tumefaciens
C;Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 01-Feb-2002
C;Accession: AG3182
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A;Reference number: AB2577; PMID:11743193
A;Accession: AG3182
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-328 <KUR>
A;Cross-references: GB:AE008667; PIDN:AA45877.1; PID:g17743620; GSPDB:GN00188
A;Experimental source: strain C58 (Dupont)
C;Genetics:
A;Gene: Atu5188
C;Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology

Query Match
Best Local Similarity 9.0%; Score 173.5; DB 2; Length 328;
Matches 94; Conservative 44; Mismatches 146; Indels 85; Gaps 16;

QY 14 GDKAVVKTDSVPELKEGTALVKEAVAGNPTDM--KHIAK-IGPESILGCDIAGTV 70
DB 11 GPEAFPLVEVPKRLPRAQGVILRVNATSIINPLDYVRDQYRDLVPLPATIGHDVSGVVE 70
QY 71 KLGPNASTDLLKVDSTFGFVHGAQOTDPKNGAFKFAEYAVYPPLYK--SNLHSTADEIS 128
DB 71 ATGPGV--TWFAQDE-----VWYTPQIFDGPGRGSAEHNVANENIGKPSLTHL----- 119
QY 129 EGVKNFESASLPVSLTTAGVSLCHLGSKMEHNPSTPOHTHPIILMGATAVGQOLIQ 188
DB 120 -----BAASILVGTAMEALVSRALRV-----GESILIHGAGAGVGHVAILQ 162
QY 189 VAKHINAYTKIYTVASKKEKLKSYGADVDVPHDAGVIEQIKSKYPL--QHVIDAVG 246
DB 163 VAAVIGA--KYVTVVEENFEARSAGADVVIDYRKEDYVAIMRTEGLGVDDVVDITIG 220
QY 247 SEDSIPEAKVTVADSLPATLLEVPMTIESIPEIRKDNKIDITLLYRASGCEILGAT 306
DB 221 GE-----ILSHPKYLAOF--GRVYSIVDIARPN-----LIEAWGR----- 255
QY 307 RFPASPEYHEATVKFYKFINPHLNGDIHNNIKVFSN-----GLDDVPALTEGI 356
DB 256 ---NASYH-----FVFTRONQKLNELNVLVERGQLRPHVGAVYSLADLP--LAHAL 302
QY 357 KEGKKNVK 365

Wed Jun 18 14:19:13 2003

us-10-081-644-2.rpr

Page 7

Db 303 LKPNNGLR 311

Search completed: June 17, 2003, 13:02:36
Job time : 29 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2003, 09:23:16 ; Search time 18 Seconds

(without alignments)
852.569 Million cell updates/sec

Title: US-10-081-644-2

Sequence: 1 MSVPTQKAVIIEGDKAVK.....ALTEGKEGKNVKKVAVRL 370

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1065.5	55.5	376	1 YNNA_YEAST	P53912 saccharomyc
2	1021.5	53.2	376	1 YL60_YEAST	P54007 saccharomyc
3	1015.5	52.9	368	1 YCZ2_YEAST	P25608 saccharomyc
4	370.5	19.3	297	1 TOXD_COCCA	P54006 cochlidoctol
5	254	13.2	309	1 OORL_ARATH	O92uc1 arabidopsis
6	175.5	9.1	348	1 ADH2_CANAL	O94038 candida alb
7	161	8.4	329	1 OOR_HUMAN	O08257 homo sapien
8	159	8.3	329	1 OOR_CAVDO	P11415 cavia porce
9	158.5	8.3	331	1 OOR_MOUSE	P47189 mus musculu
10	155	8.1	340	1 OOR_LETAM	P42665 leishmania
11	154	8.0	365	1 YM2_YEAST	P28625 saccharomyc
12	153.5	8.0	345	1 P1_ARATH	O39172 arabidopsis
13	153.5	8.0	348	1 ADH1_KLUMA	O07288 kluyveromyc
14	152.5	7.9	348	1 ADH2_PICST	O13309 pichia stip
15	149.5	7.8	348	1 ADH1_PICST	O00097 pichia stip
16	148	7.7	330	1 OOR_BOVIN	O97764 bos taurus
17	148	7.7	330	1 OOR_LAMGU	O28452 lama guanac
18	146	7.6	343	1 P2_ARATH	O39172 arabidopsis
19	145.5	7.6	350	1 ADH1_KLUMA	P20369 kluyveromyc
20	143.5	7.5	350	1 ADH1_CANAL	P43067 candida alb
21	141.5	7.4	337	1 ADH1_BACST	P12311 bacillus st
22	140.5	7.3	347	1 ADH2_KLUMA	O99462 kluyveromyc
23	140	7.3	341	1 TDH_ECOLI	P07993 escherichia
24	140	7.3	341	1 VAI1_TORCA	P19333 torpedo cal
25	134.5	7.0	374	1 ADH3_KLUMA	P49384 kluyveromyc
26	132	6.9	334	1 OOR_YEAST	P38230 saccharomyc
27	131.5	6.8	375	1 ADH3_YEAST	O06004 bacillus su
28	130.5	6.8	352	1 DHSD_BACSU	O06004 bacillus su
29	130	6.8	348	1 DHSD_BOMO	O02912 bombyx mori
30	129.5	6.7	348	1 ADH2_KLUMA	P49383 kluyveromyc
31	128	6.7	340	1 ADH1_RHIME	O31186 rhizobium m
32	127	6.6	339	1 ADH3_BACST	P42328 bacillus st
33	125.5	6.5	347	1 ADH2_YEAST	P00331 saccharomyc

34	124.5	6.5	347	1 ADH1_YEAST	P00330 saccharomyc
35	124.5	6.5	357	1 CAD4_TOBAC	P30359 nicotiana t
36	121.5	6.3	359	1 MTD_MERSA	O82515 medicago ba
37	121	6.3	399	1 DHSD_RAT	P27867 rattus norv
38	120.5	6.3	339	1 ADH2_BACST	P42327 bacillus st
39	120	6.2	365	1 MTD_APIGR	O38707 apium grave
40	118.5	6.2	351	1 ADH5_YEAST	P38113 saccharomyc
41	118	6.1	300	1 VAI1_HUMAN	O99536 homo sapien
42	117.5	6.1	367	1 ADH2_EMENT	P54202 emeticella
43	117	6.1	336	1 ADH1_ECOLI	P39451 escherichia
44	117	6.1	347	1 ADH1_SULSR	P50381 sulfobus
45	117	6.1	375	1 DHSD_MOUSE	O64442 mus musculu

ALIGNMENTS

RESULT 1	ID	YNNA_YEAST	STANDARD	PRT	376 AA.
AC	P53912	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)				
DT	01-OCT-1996 (Rel. 34, Last annotation update)				
DE	Hypothetical 41.2 kDa protein in PPR1-TOM22 intergenic region.				
GN	YNL134C OR N1214 OR N1847				
OS	Saccharomyces cerevisiae (Baker's yeast).				
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;				
OX	Saccharomycetales; Saccharomycetaceae; Saccharomycetes.				
NC	NCBI_TaxID=4932;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=S288C; PubMed=8619318;				
RA	MEDLINE=96109932; Busseu F., Jacquet M.;				
RT	"A 43.5 kb segment of yeast chromosome XIV, which contains MFA2, ME2, CAP/SRV2, NAM9, FR1/PPR1/RBP1, MOM22 and CPT1, predicts an RT adenosine deaminase gene and 14 new open reading frames."				
RL	Yeast 11:1195-1209(1995).				
CC	-1- SIMILARITY: BELONGS TO THE YCR102C/YLR460C/YNL134C FAMILY.				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@sib-sib.ch).				
DR	EMBL; Z46843; CAA86891.1; -				
DR	EMBL; Z71410; CAA86016.1; -				
DR	SGD; S0005078; YNL134C.				
DR	InterPro; IPR002085; Adh_zn_family.				
DR	Pfam; PF00107; adh_zinc_1.				
KW	Hypothetical protein.				
SO	SEQUENCE 376 AA; 41164 MW; AE39BCDA46C3B9 CRC64;				
Query Match	55.5%; Score 1065.5; DB 1; Length 376;				
Best Local Similarity	54.0%; Pred. No. 5.8e-71;				
Matches	201; Conservative 59; Mismatches 109; Indels 3; Gaps 2;				
QY	2 SVPTQKAVIIEGDKAVVTVDSVPELKEGTALVKAEAVANGPTDKHIAVYKIGPEGSLI 61				
DB	4 SIPEWKAVIENGAVVKKDIPPIPELEGEFVILIKTVAAVANGPTDKHIDFKIGPGALL 63				
QY	62 GCDIAGTVKLGPNVSTDLKVDGTFGFGVHAGASQTPKNGCAFAEYARYPLFYK--SN 118				
DB	64 GCDAGQIVKLGPNVDAARFAIGDYIYGVHAGASVAFENGCAFAEYASISSEYARYPAR 123				
QY	119 LTHSTADEISEGPKVNFESASLPLVSLTAVGSLCHHLLGSKWEHDPSTQHTHPLIIVG 178				
DB	124 FRLGCKDLPFGVSVSLGAVSLPVLSTTAGMILTHSFGLDWTWPKSKAQRDQPIIFWCG 183				

```

QY 179 ATAVGQOLIOVAKHINAYTKIVTASRKHEKLLKSYGADVDVPHDAGVIEQIKSKYPNL 238
DB 184 ATAVGQOLIOVAKHINAYTKIVTASRKHEKLLKSYGADVDVPHDAGVIEQIKSKYPNL 243
QY 239 QHVIDAVGSDSDIPEAVKVTADSLPATLLEVPMTIESIPEIRKDNVKIDITLLYPASG 298
DB 244 PYLVDVCSNMTETIOQVVKCAADLDLAVVQLVLTKEKDKEDERRQVNSIEGTLILYIGG 303
QY 299 QEILIGATRPASPEYHEATVFKFNPINLNGDIHNNIKVPSNGLDVPALTEGIRK 358
DB 304 NDVPEFGFTLLPADSEYKKAIFKIFNPINLNGDIHNNIKVPSNGLDVPALTEGIRK 363
QY 359 GKKNKVKYVARL 370
DB 364 GRNKGKLVAVL 375

RESULT 2
Y60 YEAST
P54007 STANDARD; PRT; 376 AA.
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 01-OCT-1996 (Rel. 34, Last annotation update)
DE Hypothetical 41.1 kDa protein ON CDC91-PAU4 intergenic region.
GN YLR460C OR L9122.7.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=8288c / AB972;
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,
RA Favelli A., Fulton L., Gattung S., Greco T., Kirsten J.,
RA Kucaba T., Langston L., Langston Y., Latreille P., Le T.,
RA Johnson D., Johnston L., Miller N., Nhan M., Pauley A., Peluso D.,
RA Mardis E., Meneses S., Miller N., Nhan M., Pauley A., Peluso D.,
RA Rifken L., Wohlfman P., Vaudin M., Wilson R., Waterston R.,
RA Wilcox L., Wohlfman P., Vaudin M., Wilson R., Waterston R.,
RA Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.
RL -1- SIMILARITY: BELONGS TO THE YCR102C/YLR460C/YNL134C FAMILY.
CC -1-
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U22383; AAB64723.1; -
CC SGI: S0004452; YLR460C.
CC InterPro: IPR002085; Adh_zn_family.
CC Pfam: PF00107; adh_zinc; 1.
CC Hypothetical protein.
CC SEQUENCE 376 AA; 41127 MW; 00139949423862F1 CRC64;

Query Match 53.2%; Score 1021.5; DB 1; Length 376;
Best Local Similarity 51.7%; Pred. No. 9.8e-68;
Matches 193; Conservative 67; Mismatches 110; Indels 3; Gaps 2;

```

```

QY 178 GATAVGQOLIOVAKHINAYTKIVTASRKHEKLLKSYGADVDVPHDAGVIEQIKSKYPNL 237
DB 183 GATAVGQOLIOVAKHINAYTKIVTASRKHEKLLKSYGADVDVPHDAGVIEQIKSKYPNL 242
QY 238 LQHVDAVGSSESDIPEAVKVTADSLPATLLEVPMTIESIPEIRKDNVKIDITLLYPASG 297
DB 243 ISYLVDCVANNODTLQOVVKCAADKODATTVELNLTENVKKNRRONVTIDILRLYSIG 302
QY 298 GQELILGATRPASPEYHEATVFKFNPINLNGDIHNNIKVPSNGLDVPALTEGIRK 357
DB 303 GHEVPPENITLPPADSEYKKAIFKIFNPINLNGDIHNNIKVPSNGLDVPALTEGIRK 362
QY 358 EGKNKVKYVARL 370
DB 363 YGKNKGKLVAVL 375

RESULT 3
Y62 YEAST
P25608 STANDARD; PRT; 368 AA.
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE 40.1 kDa protein in G1T1-PAU3 intergenic region.
GN YCR102C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC Grivell L.A., de Haan M., Maat M.J.;
RA Submitted (MAR-1992) to the EMBL/GenBank/DBJ databases.
RL -1- SIMILARITY: BELONGS TO THE YCR102C/YLR460C/YNL134C FAMILY.
CC -1-
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X59720; CAA42244.1; -
CC PIR: S19414; S19414.
CC SGI: S0000699; YCR102C.
CC InterPro: IPR002085; Adh_zn_family.
CC Pfam: PF00107; adh_zinc; 1.
CC SEQUENCE 368 AA; 40121 MW; 0BEE4FBAD804AF8B CRC64;

Query Match 52.9%; Score 1015.5; DB 1; Length 368;
Best Local Similarity 53.0%; Pred. No. 2.6e-67;
Matches 194; Conservative 60; Mismatches 109; Indels 3; Gaps 2;

```


Db 242 VANONTLQGVYKCAADKQDQATVVELTVLTFEENVKKNRQNVITIDTRILYSIGSHVFPFG 301
 QY 305 ATFFPSPPEHTEATKVKRINPHLNGDIHNNIKVFSGLDDVPALTEGIEGKKVAV 364
 Db 302 GTFEPADPEARRAATEFVKFIPKISDGIHHPARVYKNGLYDVPRILEDIKIKRNGSE 361
 QY 365 KYVAVL 370
 Db 362 KLVAVL 367

RESULT 4

TOXD COCCA STANDARD; PRT; 297 AA.
 AC P54006;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 TOXD protein.

OC Cochliobolus carbonum (Bipolaris zeicola).
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Dothideomycetes;
 OC Pleosporales; Pleosporaceae; Cochliobolus.
 OC NCBI_TaxId=5017;
 RN [1]
 RP SOURCE FROM N.A.
 RC STRAIN=ATCC 90305 / SB111;
 RA Cheng Y.-Q., Ann J.-H., Walton J.D.;
 RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.

CC -1- FUNCTION: UNIQUE TO ISOLATES THAT MAKE THE CYCLIC PEPTIDE HC-
 CC TOXIN, BUT HAS NO ROLE IN TOXIN BIOSYNTHESIS.
 CC -1- SIMILARITY: BELONGS TO THE YCR102C/YLR460C/YML134C FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; X92391; CAA63129.1; -
 DR InterPro: IPR002085; Adh_zn_family.
 DR Pfam: PF00107; adh_zinc_1.
 SQ SEQUENCE 297 AA; 32549 MW; 0FFB38C93E4521A1 CRC64;

Query Match 19.3%; Score 370.5; DB 1; Length 297;
 Best Local Similarity 34.0%; Pred. No. 4e-20;
 Matches 102; Conservative 53; Mismatches 112; Indels 33; Gaps 12;

QY 5 TTQKAVITIG-DAVAVKTVDSVPELKEGTLVYKVEAVAGNPTWKKIAYKIGEGSILGC 63
 Db 2 TFKKAVITAPPHARIVSDRLIPKLDYLVRTVSALNPPTWKKI-LRLSPGGLVGC 60
 QY 64 DIAGTVVKGPNASTLDLKVGDTEGFGFVHGASQTDPPKNGAFAYARVYPLFYKSNLTHST 123
 Db 61 DYAGTVEEVGRSKYKPFKKDRVCGRAHGNAVPSDDGTAEVITV-----KGDIDAMI 114
 QY 124 ADEISBGVKNFESASLPSVSLTTAGVSLCHLGSCKMEWHPSTP-OHTHEPLTWGATNV 182
 Db 115 PENLST-----FGEAATLIGVIGITVGGLVQSL-KLSW-PTTPIEHAVALILYGSSTAI 165
 QY 183 GQDLQVAKHINAYTKIVYASKKEKLSYGADVDVFDYHDGAVLEQISKTPN-LQHY 241
 Db 166 GTLAIQIAK-LSGY-RITTTCSPHHELMKSLGADLVFDYHEITSDHTRCTQNTKIKYV 223
 QY 242 IDAVGSDS--PEAVYVTAADSLPATLLEVPMTTISPEIRKDNVKKIDITLLYASGQ 299
 Db 224 FDTISIDVSAKFCDPMSTGEYSALVDV-----SLAKTINISMTLAIYVVLGE 273

RESULT 5

QOOL ARATH STANDARD; PRT; 309 AA.
 AC O9ZUC1;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Quinine oxidoreductase-like protein At1g23740.
 GN AT1G23740 OR F508.29.

OS Arabidopsis thaliana (mouse-ear cress).
 OC Eukaryota; Viridiplanetae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OC NCBI_TaxId=3702;
 RN [1]
 RP SOURCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=21016719; PubMed=11130712;
 RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
 RA White O., Alonso J., Altieri H., Araujo R., Bowman C.L., Brooks S.Y.,
 RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
 RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
 RA Dunn P., Egu P., Feldblum T.V., Feng J.-D., Feng B., Fujii C.Y.,
 RA Gali J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Hultzer L.,
 RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
 RA Kim C.J., Koo H.L., Kremetska I., Kuritz D.B., Kwan A.J., Lam B.,
 RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
 RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Malt R., Marshall A.,
 RA Millscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
 RA Palto G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
 RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
 RA Sun H., Tallon L.J., Tamblinga G., Toriumi M.J., Town C.D.,
 RA Utterback T., Van Aken S., Vayberg M., Vysotskaia V.S., Walker M.,
 RA Wu D., Xu G., Frazer C.M., Venter J.C., Davis R.W.,
 RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
 RT thaliana".
 RL Nature 408:816-820(2000).

CC -1- SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE
 CC FAMILY. QUINONE OXIDOREDUCTASE SUBFAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; AC005990; AAC98029.1; -
 DR SWISS-2DPAGE; O9ZUC1; ARATH.
 DR InterPro: IPR002085; Adh_zn_family.
 DR InterPro: IPR002364; QOR_zeta_crystal.
 DR Pfam: PF00107; adh_zinc_1.
 DR PROSITE; PS01162; QOR_ZETA_CRYSTAL; FALSE_NEG.
 SQ SEQUENCE 309 AA; 32775 MW; EC328042771BEE6 CRC64;

Query Match 13.2%; Score 254; DB 1; Length 309;
 Best Local Similarity 32.8%; Pred. No. 1.5e-11;
 Matches 84; Conservative 35; Mismatches 97; Indels 40; Gaps 10;

QY 15 DRAVTVDSVPELKEGTLVYKVEAVAGNPTWKKIAYKIGEGSILGCDIAGTV 70
 Db 13 DVLKESNIVPEIKEDQYLIVVAALNPVAKRQGFKAIDSLPLFVPGDVAQVY 72
 QY 71 KLGPNASTLDLKVGDTEGFGFVHGASQTDPPKN-GAFAYARVYPL-FYKSNLTHSTADEI 127
 Db 73 KYG-SAVKDLKEGDEVYANVSEKALEGPKQFSLAEYTAVERKLTALKPKNI----- 123
 QY 128 SEGVPVKNFESASLPSVSLTTAGVSLCHLGSCKMEWHPSTPQTHPLLLINGATVAGQQL 187
 Db 124 -----DFAQAGLPLALETADGVL-----RTEPSAG-----KSLTVLNGAGGVSLVI 167
 QY 188 QVAKHINAYTKIVYASKKEKLSYGADVDVFDYHDGAVLEQISKYVNLGHVIDAVGS 247

Db 168 QLAHVVGASKVAATASTKLELVRSIGADLAIDYTKEN-1EDLPDKY---DVVFDALIGM 223
 QY 248 EDS-----IPEAKYKTA 259
 Db 224 CDKAVKVIKESGKVA 239

RESULT 6

ID ADH2 CANAL STANDARD; PRT; 348 AA.
 AC 094038;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Alcohol dehydrogenase 2 (EC 1.1.1.1).
 GN ADH2-OR CA41C10.04.
 OS Candida albicans (Yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; mitosporic Saccharomycetales; Candida.
 NCBI_TaxID=5476;
 [1]

SEQUENCE FROM N.A.

RC STRAIN=1161;
 RA Taylor K., Harris D., Barrell B.G., Rajandream M.A.;
 RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- CATALYTIC ACTIVITY: An alcohol + NAD(+) = an aldehyde or ketone +
 NADH.
 CC -1- COFACTOR: ZINC (By similarity).
 CC -1- SUBUNIT: HOMOTETRAMER (By similarity).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
 CC -1- SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE
 FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL, AL033501, CA21988.1;
 DR InterPro: IPR002328; ADH_zinc.
 DR InterPro: IPR002085; Adh_zn_family.
 DR Pfam: PF00107; adh_zinc.1.
 DR PROSITE: PS00059; ADH_ZINC.1.

CC Oxidoreductase; Zinc; NAD.
 METAL 44 44 ZINC (CATALYTIC) (BY SIMILARITY).
 METAL 67 67 ZINC (CATALYTIC) (BY SIMILARITY).
 FT METAL 98 98 ZINC (SECOND ATOM) (BY SIMILARITY).
 FT METAL 101 101 ZINC (SECOND ATOM) (BY SIMILARITY).
 FT METAL 104 104 ZINC (SECOND ATOM) (BY SIMILARITY).
 FT METAL 112 112 ZINC (SECOND ATOM) (BY SIMILARITY).
 FT METAL 154 154 ZINC (CATALYTIC) (BY SIMILARITY).
 SQ SEQUENCE 348 AA; 36607 MW; 3D6E6738522ADCE CRC64;

Query Match 9.1%; Score 175.5; DB 1; Length 348;
 Best Local Similarity 24.6%; Pred. No. 1e-05;
 Matches 98; Conservative 49; Mismatches 138; Indels 113; Gaps 19;

QY 1 MSVPTQKAVIIE-GDKAVVKTVSVPELKEGALYKVEAVAGNPTD--WK--HIAY 52
 Db 1 MSVPTQKAVIIEFTNGKLEBYK-DIPVKKPKNELLINVKXSGVCHTDLHAMKGDWPLAT 59
 QY 53 KI-----GREGSILGCDIAGTVYKLGPNASTDLKXGD-TGPFVYGA-----SQDPK 99
 Db 60 KLPLVGGHEG-----AGVVVALGENVK-GWKYGDVAGVWKLGSCLNCEYCOGSAEPN 111
 QY 100 -----NGAFAYRAYRPPPLFYKSNLTHSTADEISEGPKNFESASLPSLTUA 148
 Db 112 CAEADLSGYTHDGSFQGYA-----TADAV-----QARITAGDTLA 147

QY 149 GVS-LCHHLGSMENHPSTPQHTPLLIWGATAVAGQOLIQAHIINAYTKIVTASKK 206
 Db 148 NVAPILCAGTVYKALKTAELKAGVVAISGAAGLISLAVYAKAM-GYRVLAIDGDE 206
 QY 207 HEKLKSYGADVDFY-HDAGVIEQIKSKYKPNQH-VIDAVSGEDSIPKAYKTADSLA 264
 Db 207 KGEFVKSLEGETPDTPEKEDVAVKATNGSPHGVINVSERAIQGS-----256
 QY 265 TLEAVPMITIESIPERIKDNVXIDITLLYRASGGEILGATFPA-----SPEYHATVK 320
 Db 257 -----TEYRTLGKVLVG---LPGAKISTVDPVAVIK 287
 QY 321 FVKFIPHLNNGDIHHNNTKVSNGDDVDPALTEGKE 358
 Db 288 TIOIKSGYVGNRMDTAADVFTFRLIKCPKIVGISE 325

RESULT 7

QY QOR_HUMAN STANDARD; PRT; 329 AA.
 ID QOR_HUMAN
 AC 008257;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Quinone oxidoreductase (EC 1.6.5.5) (NADPH:quinone reductase) (Zeta-
 DE crystallin).
 GN CRYZ.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NCBI_TaxID=9606;
 [1]

SEQUENCE FROM N.A.

RC TISSUE=Liver;
 RX MEDLINE=93221534; PubMed=8466529;
 RA Gonzalez P., Rao P.V., Zigler J.S. Jr.;
 RL "Molecular cloning and sequencing of zeta-crystallin/quinone
 RL reductase cDNA from human liver."
 RL Biochem. Biophys. Res. Commun. 191:902-907 (1993).
 RN [2]

SEQUENCE FROM N.A.

RC TISSUE=Liver;
 RX MEDLINE=94375054; PubMed=8088825;
 RA Gonzalez P., Rao P.V., Zigler J.S. Jr.;
 RL "Organization of the human zeta-crystallin/quinone reductase gene
 RL (CRYZ)."
 RL Genomics 21:317-324 (1994).
 CC -1- FUNCTION: DOES NOT HAVE ALCOHOL DEHYDROGENASE ACTIVITY. BINDS
 CC NADP AND ACTS THROUGH A ONE-ELECTRON TRANSFER PROCESS.
 CC ORTHOQUINONES ARE THE BEST SUBSTRATES. MAY ACT IN THE
 CC DETOXIFICATION OF XENOBIOTICS.
 CC -1- CATALYTIC ACTIVITY: NADPH + quinone = NADP(+) + semiquinone.
 CC -1- SUBUNIT: HOMOTETRAMER.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- TISSUE SPECIFICITY: ONLY VERY LOW AMOUNTS IN THE LENS.
 CC -1- SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE
 CC FAMILY. QUINONE OXIDOREDUCTASE SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL, L33278; AAA36536.1;
 DR EMBL, L31526; AAK40311.1;
 DR EMBL, L31521; AAK40311.1; JOINED.
 DR EMBL, L31522; AAK40311.1; JOINED.
 DR EMBL, L31523; AAK40311.1; JOINED.
 DR EMBL, L31524; AAK40311.1; JOINED.
 DR EMBL, L31525; AAK40311.1; JOINED.

DR HSSP; P28304; 100R.
 DR Genew; HGNC:2419; CRYZ.
 DR MIM; 123691; -
 DR InterPro; IPR002085; Adh_zn_family.
 DR InterPro; IPR002364; QOR_zeta_crystal.
 DR Pfam; PF00107; adh_zinc; 1.
 DR PROSITE; PS01162; QOR_ZETA_CRYSTAL; 1.
 DR OXIDOREDUCTASE; NADP_Zinc
 KM SEQUENCE 329 AA; 35206 MW; 68C1828911486D4E CRC64;
 SO
 Query Match 8.4%; Score 161; DB 1; Length 329;
 Best Local Similarity 24.8%; Pred. No. 0.00015;
 Matches 73; Conservative 50; Mismatches 95; Indels 76; Gaps 14;
 QY 19 VKTDVSVPELKEGTALVKEAVAGNPDMKHLAYKIGREGSL-----GCDIAGTVYKLP 74
 DB 24 LNSDVAIPKQHVLIKVAACGNPVE--TYRSGTYTRIRPLPYTPGSDVAGVSEISN 82
 QY 75 NASTDLKVDGTGFGVHGASQTPDKNGAFAYARVYPLPYFYSNLTSTADEISSEGPVK- 133
 DB 83 NNSA-FKGDGRV-----TSTISGVAEYALADHIVYKL-----PEKL 121
 QY 134 NFEASASLPVSLTTAGVSLCHLGSKMEWHSTPQHTPLLIWGATAVGQOLIYAKHI 193
 DB 122 DFGQAAIGIPYFTAYRALHSAVCVA-----GESLVHGASGVGLAACQIAR-- 170
 QY 194 NAY-TKIV-TVASKKHEKLKSYGADVDPDYHDAGVIEQIKSKYRNLOHVIDAVSEDSI 251
 DB 171 -AYGLKVLGTAGTEGQKVLONGAHEVFNRHVNIDIKI-KY-----VGEKG-- 217
 QY 252 PEAYKVTADSLPATLLEVPMTIESIPEIRKDNVKIDITLLYRASGOEILGA 305
 DB 218 -----IDIIEMLANVNSDKDSLH-SGGRVIIVGS 248
 RESULT 8
 QOR CAVPO STANDARD; PRT; 329 AA.
 ID QOR CAVPO STANDARD; PRT; 329 AA.
 AC P1415;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-OCT-1989 (Rel. 12, Last annotation update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Quinone oxidoreductase (EC 1.6.5.5) (NADPH:quinone reductase) (Zeta-crystallin).
 GN CRYZ.
 OS Cavia porcellus (Guinea pig).
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Hystriognathi; Caviidae; Cavia.
 NCBI_TaxID=10141;
 RN [1]
 SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RX MEDLINE=89378748; PubMed=2777081;
 RA Rodokanaki A., Holmes R.K., Borras T.;
 RT "Zeta-crystallin, a novel protein from the guinea pig lens is related to alcohol dehydrogenases";
 RL Gene 78:215-224 (1989).
 RN [2]
 CHARACTERIZATION.
 RC TISSUE=Liver;
 RX MEDLINE=93041895; PubMed=1420281;
 RA Rao P.V., Ziegler J.S. Jr.;
 RT "Purification and characterization of zeta-crystallin/quinone reductase from guinea pig liver";
 RL Biochim. Biophys. Acta 1117:315-320 (1992).
 RN [3]
 CHARACTERIZATION.
 RC TISSUE=Liver;
 RX MEDLINE=92112732; PubMed=1370456;
 RA Rao P.V., Krishna C.M., Ziegler J.S. Jr.;
 RT "Identification and characterization of the enzymatic activity of zeta-crystallin from guinea pig lens. A novel NADPH:quinone oxidoreductase";
 RL J. Biol. Chem. 267:96-102 (1992).

CC -1- FUNCTION: DOES NOT HAVE ALCOHOL DEHYDROGENASE ACTIVITY. BINDS
 CC NADP AND ACTS THROUGH A ONE-ELECTRON TRANSFER PROCESS.
 CC ORTHOQUINONES ARE THE BEST SUBSTRATES. MAY ACT IN THE
 CC DETOXIFICATION OF XENOBIOTICS.
 CC -1- CATALYTIC ACTIVITY: NADPH + quinone = NADP(+) + semiquinone.
 CC -1- SUBUNIT: HOMOTETRAMER.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- TISSUE SPECIFICITY: IN GUINEA PIGS IT CONSTITUTES ABOUT 10% OF THE
 CC WATER SOLUBLE PROTEINS OF THE LENS. THE DELETION OF 34 AMINO ACIDS
 CC -1- DISEASE: A GENOMIC MUTATION CAUSING THE DELETION OF 34 AMINO ACIDS
 CC WAS CLEARLY ASSOCIATED WITH A HEREDITARY NUCLEAR CATARACT IN A
 CC LINE OF STRAIN 13 GUINEA PIGS.
 CC -1- SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE
 CC FAMILY. QUINONE OXIDOREDUCTASE SUBFAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL; M26936; AAA37035.1; -
 CC PIR; J50230; CYGPZ.
 DR HSSP; P28304; 100R.
 DR InterPro; IPR002085; Adh_zn_family.
 DR InterPro; IPR002364; QOR_zeta_crystal.
 DR Pfam; PF00107; adh_zinc; 1.
 DR PROSITE; PS01162; QOR_ZETA_CRYSTAL; 1.
 DR OXIDOREDUCTASE; NADP_Zinc; Eye lens protein.
 KM SEQUENCE 329 AA; 35202 MW; 1463632C672CC34F CRC64;
 SO
 Query Match 8.3%; Score 159; DB 1; Length 329;
 Best Local Similarity 24.2%; Pred. No. 0.00015;
 Matches 71; Conservative 50; Mismatches 96; Indels 76; Gaps 14;
 QY 19 VKTDVSVPELKEGTALVKEAVAGNPDMKHLAYKIGREGSL-----GCDIAGTVYKLP 74
 DB 24 VSDVAIPKQHVLIKVAACGNPVE--TYRSGTYTRIRPLPYTPGSDVAGVSEISN 82
 QY 75 NASTDLKVDGTGFGVHGASQTPDKNGAFAYARVYPLPYFYSNLTSTADEISSEGPVK- 133
 DB 83 DVSA-FKGDGRV-----TSTISGVAEYALADHIVYKL-----PEKL 121
 QY 134 NFEASASLPVSLTTAGVSLCHLGSKMEWHSTPQHTPLLIWGATAVGQOLIYAKHI 193
 DB 122 DFGQAAIGIPYFTA-----CRAL-----FHSARARAGESVLVHGASGVGLAACQIAR-- 170
 QY 194 NAY-TKIV-TVASKKHEKLKSYGADVDPDYHDAGVIEQIKSKYRNLOHVIDAVSEDSI 251
 DB 171 -AYGLKVLGTAGTEGQKVLONGAHEVFNRHVNIDIKI-KY-----SIEKG-- 217
 QY 252 PEAYKVTADSLPATLLEVPMTIESIPEIRKDNVKIDITLLYRASGOEILGA 304
 DB 218 -----VDVIEMLANVNSDKKL-SGGRVIIVG 247
 RESULT 9
 QOR MOUSE STANDARD; PRT; 331 AA.
 ID QOR MOUSE STANDARD; PRT; 331 AA.
 AC P47199; O99L63; O62508;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Quinone oxidoreductase (EC 1.6.5.5) (NADPH:quinone reductase) (Zeta-crystallin).
 GN CRYZ.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]

RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=94224126; PubMed=8170370;
RA Gonzalez P., Hernandez-Calzadilla C., Rao P.V., Rodriguez I.R.,
RT Ziegler J.S., Jr., Borras T.;
RT "Comparative analysis of the zeta-crystallin/quinone reductase gene
in guinea pig and mouse."
RL Mol. Biol. Evol. 11:305-315(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Breast tumor;
RA Strausberg R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 130-331 FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Brain cortex;
RX MEDLINE=96216731; PubMed=8645260;
RA Kajiwara K., Nagasawa H., Shimizu-Nishikawa K., Ookura T., Kimura M.,
RA Sugaya E.;
"molecular characterization of seizure-related genes isolated by
differential screening."
Biochem. Biophys. Res. Commun. 219:795-799(1996).
CC -1- FUNCTION: DOES NOT HAVE ALCOHOL DEHYDROGENASE ACTIVITY. BINDS
NADP AND ACTS THROUGH A ONE-ELECTRON TRANSFER PROCESS.
CC ORTHOQUINONES ARE THE BEST SUBSTRATES. MAY ACT IN THE
CC DETOXIFICATION OF XENOBIOTICS (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: NADPH + quinone = NADP(+) + semiquinone.
CC -1- SUBUNIT: HOMOTETRAMER.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE
CC FAMILY. QUINONE OXIDOREDUCTASE SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: S70056; AAB30620.2; -
DR EMBL: BC003800; AAH03800.1; -
DR EMBL: D78646; BA011463.1; -
DR HSSP: P28304; 1QOR.
DR MGD: MGI:88527; Cryz.
DR Interpro: IPR002085; Adh zn family.
DR Interpro: IPR002364; QOR_zeta_crystal.
DR Pfam: PF00107; adh_zinc; 2.
DR PROSITE: PS01162; QOR_ZETA_CRYSTAL; 1.
DR OXIDOREDUCTASE; NADP; ZINC.
FT CONFLICT 58 58 A -> T (IN REF. 2).
FT CONFLICT 131 133 IPY -> TMD (IN REF. 3).
SQ SEQUENCE 331 AA; 35268 MW; 35816C043EF16A2 CRC64;

Query Match 8.3%; Score 158.5; DB 1; Length 331;
Best Local Similarity 22.4%; Pred. No. 0.00011;
Matches 72; Conservative 53; Mismatches 119; Indels 77; Gaps 13;

QY 19 VKTIDVSPBELKEGATLVKVAVAGNPTDM--KHIAVYKIGPGSL--GCDIAGTVVTKGN 75
DB 24 LQSVVVVPQSHOVLTKVHACGVNPEYTYIRSGAYSRKPLPTTPGSDVAGLIESGDK 83
QY 76 ASTDLKVGDTGFGFVHGASQTPKNGAFAEYA-----RVYPPLEFYKSNLTHSTADEISEG 130
DB 84 VSA-FKSGDRVFC-----STVSGGYAFELALADTIYP-----LPE- 119
QY 131 PVKNFESAASLPVSLTTAGVSLCHHLGSKMEWEPSTQHTHPLLIWGAATVAGQOLIQVA 190
DB 120 -TLNFRGAAALGIIFYTA-----CRAL-----FHSARARAGSEVAVHGASGGVGLATCOIA 169
QY 191 KHINAYKIYVAVAKKHEKLLKSYGADVDVDPYHAGVIEQKSKYRPLQHVIVAVGSEDS 250
DB 170 R-AHGLKVLGTAGSEBGRKVLQNGAHEVFNHEKANTYDIKIM-----SVGDKD 218

QY 251 IPEAVKVTADSLPATILEVPMPTIESIPEIRKDNKYDITLLYRASGOELLIGATRPFA 310
DB 219 -----GVDTIEMLANENISNDKLK--SHGRVYVVGGRGI 254
QY 311 SPEYHEATVYKFEINPHLNN 331
DB 255 EINPRDTMAKETSIIIGVSLSS 275

RESULT 10
QOR_LEIAM
ID QOR_LEIAM STANDARD; PRT; 340 AA.
AC P42865;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Possible quinone oxidoreductase (EC 1.6.5.5) (NADPH:quinone
reductase) (P36).
OS Leishmania amazonensis.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxId=5659;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95107343; PubMed=7808470;
RA Liu X., Chang K.P.;
RT "Identification by extrachromosomal amplification and overexpression
of a zeta-crystallin/NADPH-oxidoreductase homologue constitutively
expressed in Leishmania spp.;"
RL Mol. Biochem. Parasitol. 66:201-210(1994).
CC -1- CATALYTIC ACTIVITY: NADPH + quinone = NADP(+) + semiquinone.
CC -1- SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE
CC FAMILY. QUINONE OXIDOREDUCTASE SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L11705; AAA73554.1; -
DR Interpro: IPR002085; Adh zn family.
DR Interpro: IPR002364; QOR_zeta_crystal.
DR Pfam: PF00107; adh_zinc; 1.
DR PROSITE: PS01162; QOR_ZETA_CRYSTAL; 1.
DR OXIDOREDUCTASE; NADP; ZINC.
SQ SEQUENCE 340 AA; 36272 MW; 4D9255AB56856625 CRC64;

Query Match 8.1%; Score 155; DB 1; Length 340;
Best Local Similarity 23.5%; Pred. No. 0.00011;
Matches 92; Conservative 58; Mismatches 151; Indels 90; Gaps 19;

QY 1 MSVPTQKAVIIBGDAVYKTDVSV-----PELKEGATLVKVAVAGNPTDMKHIAVYK 54
DB 1 MSSPSPFKKQLQVSLSKDRSSFTSVSEALPEEVPBGWVSVKXAGVNASDLNF----- 55
QY 55 GPGGSL-----GCDIAGTVVTLGPNASTDLVGPDTGFGFVHGASQTPKNGAFAP 105
DB 56 -TNGSFPKAVQPPFDGCFEAGTIVDIGVA-NVAVGD-----HVTLM-----QYCFPE 104
QY 106 YARVYPPLEFYKSNLTHSTADEISEG--PVKNFESAAS-LPVSLTTAGVSLCHHLGSKMEW 162
DB 105 F-----LDAPARERCIPBELKPEYSVLPVSLTAVAL-GEVGRVKG 146
QY 163 HESTPQHTHPLLIWGAATVAGQOLIQVAKHINAYKIYVAVAKKHEKLLKSYGADVDVDPY 222
DB 147 DVA-----LVTAAAGGQIAVQLKRVYGCCTYIGCSSSEKAEFLKSIIGCDHVINY 198
QY 223 HDAGVIEQKSKYP-NLQHVIVAVGSEDSIPEAYKVTADSLPATL-----EVVPM 272
DB 199 KTESLDGRHLCEPKGVADVVECVGH-TFNDVAVHVAVHARVVIIGSISYSKSGEVVFP 257

QY 273 TIESIEIRKONKVIDITLLVR-ASGOELLGATFPASPEXHEATVFKFINPHLNN 331
 DB 258 SDPS-----GTSTMLLVKASASLNGPFL-----PQPHDVIPTKMANLLOYLKA 301
 QY 332 GDH-HMNIVKPSNGLDVDPALTEGIGKEGKN 361
 DB 302 GQVKLFPVKKVF-HGLSSVADAVDHLYSGAN 331

RESULT 11
 YM27 YEAST STANDARD; PRT; 365 AA.
 AC P28625;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE Hypothetical 41.6 kDa protein in IMP1-HLJ1 intergenic region (RF1095).
 YMR152W OR YMR9375.22 OR YMR520.01.
 Saccharomyces cerevisiae (Baker's Yeast).
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 NCBI_TaxID=4932;
 [1] SEQUENCE FROM N.A.
 RX MEDLINE=91360060; PubMed=186606;
 RA Behrens M., Michaelis G., Pratie E.;
 RT "Mitochondrial inner membrane protease 1 of Saccharomyces cerevisiae
 RT show sequence similarity to the Escherichia coli leader peptidase.";
 RL Mol. Genet. 228:167-176(1991).
 RN [2]
 RP SEQUENCE OF 1-316 FROM N.A.
 RC STRAIN=S288c / AB972;
 RA Badcock K., Churcher C., Barrell B.G., Rajandream M.A., Walsh S.V.;
 RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 196-365 FROM N.A.
 RC STRAIN=S288c / AB972;
 RA Hunt S., Bowman S., Barrell B.G., Rajandream M.A.;
 RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: SOME, TO YEAST AST1/AST2.
 CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; S55518; AAB19702.1; -;
 DR EMBL; Z47071; CAAB7367.1; -;
 DR EMBL; Z49705; CAAB9788.1; -;
 DR PIR; S16819; S16819.
 DR SGD; S0004760; YMR152W.
 DR InterPro: IPR002085; Adh_zn_family.
 DR Pfam: PF00107; adh_zinc_1.
 KM Hypothetical protein; Transmembrane.
 FT TRANSMEM 140 158
 FT POTENTIAL.
 FT CONFLICT 121 121 V -> I (IN REF. 1).
 FT CONFLICT 234 234 S -> I (IN REF. 1).
 FT CONFLICT 280 280 D -> N (IN REF. 1).
 FT CONFLICT 296 296 L -> S (IN REF. 1).
 FT SEQUENCE 365 AA; 41637 MW; 88F6453D9E18A16 CRC64;

Query Match 8.0%; Score 154; DB 1; Length 365;
 Best Local Similarity 23.5%; Pred. No. 0.00041;
 Matches 63; Conservative 49; Mismatches 116; Indels 40; Gaps 12;

QY 6 TQAVIIEGKAVKTVDSVPEIK-----ETALVKYEAANGPTDW-KH---IAYKIGPE 57
 DB 7 TNKSVTVVNNVTPTITSELDRSCQDDEVIEVHAALNINIDFTYHQLCNSTYFGKY 66

QY 58 GSILGDIAGTVKLGPNASTDLKYVDGTGFGF---VHGASQTDPKNGAFAYARVYP--- 111
 DB 67 PKTYSNDYSGVITIKAKQVDNRMKGDKNKMYSHYIG-----EKGITLTHLILNPAD 120
 QY 112 -PLFYKSNLTHSTADEISEGPVKNFESAASLPVSLTTAGVSLCHLIGSKJEMWHPSTPQHT 170
 DB 121 VPL---THWVEVPKDE--NDPYDDFFVYAAWPLTFTASTL---YDFKKDWTSDS----- 168
 QY 171 HPLLWGGATVAGQOLLOVAKHINATKYVTYASKKHETLLSYGADVDFDHDAGVITQ 230
 DB 169 -KVLVIGASTSVSYAFVHIATKNYFNIGTVVGCISKNSIRNNKLTGYDLVLPYDEGSIVEN 227
 QY 231 IKS-KYPNLOH-----VIDAVGSEDSIP 252
 DB 228 VKKLKQSVLENDKPFMIIPDSVGNHDFP 255

RESULT 12
 P1 ARATH
 ID P1 ARATH STANDARD; PRT; 345 AA.
 AC Q35172;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Probable NADP-dependent oxidoreductase P1 (EC 1.-.-.-).
 GN P1 OR AT5G16970 OR FK13.120.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC euroside II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=96064691; PubMed=7592828;
 RA Babiychuk E., Kushnir S., Belles-Boix E., van Montagu M., Inze D.;
 RT "Arabidopsis thaliana NADPH oxidoreductase homologs confer tolerance
 RT of yeasts toward the thiol-oxidizing drug diamide.";
 RL J. Biol. Chem. 270:26224-26231(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=21016721; PubMed=1130714;
 RA Tabata S., Kaneko T., Nakamura Y., Koriari H., Kato T., Asamizu E.,
 RA Miyajima N., Sasamoto S., Kimura T., Hosouchi T., Kawashina K.,
 RA Kohara M., Matsumoto M., Matsuno A., Muraki A., Nakayama S.,
 RA Nakazaki N., Naito K., Okumura S., Shinjo S., Takeuchi C., Wada T.,
 RA Watanabe A., Yamada M., Yasuda M., Sato S., de la Bastide M.,
 RA Huang E., Spiegel L., Gnoj L., O'Shaughnessy A., Preston R.,
 RA Habermann K., Murray J., Johnson D., Rohlffing T., Nelson J.,
 RA Stonking T., Pepin K., Spieth J., Sekhon M., Armstrong J., Becker M.,
 RA Belter E., Cordum H., Cordes M., Courtney L., Courtney W., Dante M.,
 RA Du H., Edwards J., Fryman J., Haakensen B., Lamar E., Latreille P.,
 RA Leonard S., Meyer R., Mulvaney E., Ozeresky P., Riley A., Strommatt C.,
 RA Wagner-McPherson C., Wollam A., Yeakum M., Bell M., Dedhia N.,
 RA Barnell L., Shah R., Rodriguez M., Hoon See L., Vil D., Baker J.,
 RA Kirchoff K., Toth K., King U., Bahrec A., Miller B., Merra M.,
 RA Martensen R., McCombie W.R., Wilson R.K., Murphy G., Bancroft I.,
 RA Volckaert G., Wambutt R., Duesterhoeft A., Stiekema W., Pohl T.,
 RA Eutlian K.-D., Terry N., Hartley N., Bent E., Johnson S.,
 RA Landham S.-A., McCullagh B., Robben J., Grymonprez B., Zimmermann W.,
 RA Rampeger U., Wedler H., Balke K., Wedler E., Peters S.,
 RA van Staveren T., Dirkse W., Moolman P., Klein Lankhorst R.,
 RA Weltensegger M., Botne G., Rose M., Hauf J., Bernsener S., Hempel S.,
 RA Feldpausch M., Lamberth S., Villarroel R., Gietlen J., Ardiles W.,
 RA Bente O., Lemcke K., Kolesov G., Mayer K.F.X., Rudd S., Schöof H.,
 RA Schueller C., Zaccaria P., Mewes H.-W., Bevan M., Franz P.F.;
 RT "Sequence and analysis of chromosome 5 of the plant Arabidopsis
 RT thaliana.";
 RL Nature 408:823-826(2000).
 CC -1- SIMILARITY: MAY PLAY A DISTINCT ROLE IN PLANT ANTIOXIDANT DEFENSE
 CC AND IS POSSIBLY INVOLVED IN NAD(P)/NAD(P)H HOMEOSTASIS.
 CC -1- SIMILARITY: BELONGS TO THE L4BD FAMILY OF NADP-DEPENDENT

OXIDOREDUCTASE.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; Z49768; CAA89838.1; -
 DR EMBL; AL391141; CAC01710.1; -
 DR InterPro; IPR02085; Adh_zn_family.
 DR Pfam; PF00107; adh_zinc; 1.
 KM Oxidoreductase; NADP.
 SQ SEQUENCE 345 AA; 38134 MW; 5AFCBB2948B2680 CRC64;

Query Match 8.0%; Score 153.5; DB 1; Length 345;

Best Local Similarity 22.5%; Pred. No. 0.00041;
 Matches 91; Conservative 67; Mismatches 142; Indels 105; Gaps 21;

6 TQKAVII-----EGDKAVKTDVSVPELKEGT--ALYKVAVAGNPDMKHIAKI 54
 4 TNKQVILKDYVSGPTSDDFDTTTLVEL-RVPEGTNSVLKMLYLSCDP---YMKIRM 58
 55 G-PGSGILGCDIAGT-----VVKLGPNASTDLKVDGTGFVHGASQTDPKNGAFA 104
 59 GKRPSTPALAQAATTPOGPIGVSRISGSHDVKKGDLLMKIV-----AME 107
 105 EYARVYPLPYKSNLTHSTADEISEGPKVKNESAASLPVSLTTAGVSLCHLHSGKMEHP 164
 108 EYVITMTTHAFKIOH-----TDVPLSYTGLHGFMTAVAGF-----YEV 150
 165 STPQHTPLIWMGATVAGQOLIOVAKHINAYTKIVAKSKHKILKS-YGADVVDY- 222
 151 CSPSEGTTVVAASAAGVQLVGLAKMGCTV-VGSGSKKVDLTKTFGPDARFYK 209
 223 HDAGVIEQIKSKYPN-LQHVDAVGSSEDSIPEAYKVTADSLPATLLEVVPWTIESIBEI 281
 210 EESDLTAAKRCFPNGIDYFENVGK-----MDAVLVNM----- 245
 282 RKDNVKIDITLL-----YRASGE-----TLGATR--FPASPEYHEATVFKVFIN 326
 246 ---NMHGRIAVCGMISQYNLENQGVNLSNIIYKIRIQGFVVS-FYDKYSKLEFVL 301
 327 PHLNGDIIHNNIKVFSNGLDVDPALTEGIEKGN--KNVYKVAR 369
 302 PHIREKITTY--VEDVADGEKAPBALVGLFHGKNGKQVVVAR 344

RESULT 13

ADH1_KLUMA

ID ADH1_KLUMA STANDARD; PRT; 348 AA.

AC Q07288;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Alcohol dehydrogenase I (EC 1.1.1.1).

OS ADH1.
 OS Kluyveromyces marxianus (Yeast) (Kluyveromyces fragilis).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
 ON NCBI_TaxID=4911;
 RX MEDLINE=93250057; PubMed=8485163;

RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 12424;
 RA Ladrerie J.M., Delcour J., Vandenhaute J.;
 RT "Sequence of a gene coding for a cytoplasmic alcohol dehydrogenase
 from Kluyveromyces marxianus ATCC 12424.";
 RL Biochim. Biophys. Acta 1173:99-101 (1993).
 CC -1- CATALYTIC ACTIVITY: An alcohol + NAD(+) = an aldehyde or ketone +
 NADH.

CC -1- COFACTOR: ZINC.
 CC -1- SUBUNIT: HOMOTETRAMER.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE
 CC FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; X60224; CAA42785.1; -
 DR PIR; S32521; S32521.
 DR InterPro; IPR00328; ADH_zinc.
 DR InterPro; IPR02085; Adh_zn_family.
 DR Pfam; PF00107; adh_zinc; 1.
 DR PROSITE; PS00059; ADH_ZINC; 1.

KM Oxidoreductase; zinc; NAD; Multigene family.
 FT METAL 44 44 ZINC (CATALYTIC) (BY SIMILARITY).
 FT METAL 67 67 ZINC (CATALYTIC) (BY SIMILARITY).
 FT METAL 98 98 ZINC (SECOND ATOM) (BY SIMILARITY).
 FT METAL 101 101 ZINC (SECOND ATOM) (BY SIMILARITY).
 FT METAL 104 104 ZINC (SECOND ATOM) (BY SIMILARITY).
 FT METAL 112 112 ZINC (SECOND ATOM) (BY SIMILARITY).
 FT METAL 154 154 ZINC (CATALYTIC) (BY SIMILARITY).
 SQ SEQUENCE 348 AA; 37158 MW; A75D2EBB82E355BD CRC64;

Query Match 8.0%; Score 153.5; DB 1; Length 348;

Best Local Similarity 22.6%; Pred. No. 0.00042;
 Matches 90; Conservative 53; Mismatches 142; Indels 113; Gaps 20;

1 MSVPTQKAVII--EGDKAVKTDVSVPELKEGTALVKE-----AVAGN-PTDM 47
 1 MAIPETQGVIFYEHGSELQYK-DIVPKKPEMBELINVYSGVCHTDLHAMQGDWPLDT 59
 48 KHLIAKIGPESILGCDIAGTAVKLGPNASTDLKVD-TSGFVHGASQT----- 96
 60 K-LPIVGHGHE-----AGIVAMGENV-TGMEIGVYAGIKMLNGSCMSCECELSNEP 110
 97 -DK-----NGAEVYKRVPLPYKSNLTHSTADEISEGPV-KNFESAASLPVSLT 146
 111 NCPKADLSGYTHDSFQOYA-----TADAVQAARIPQNDVLAEPAP--L 153
 147 TAGVSLCHHLSGSKK---EWHPSPTQHTPLIWMGATVAGQOLIOVAKHINAYTKIVTV 202
 154 CAGVTYVKAALKSAHIKAGDW-----VAISGACGGISLAIOYAKM-GYRVLGID 202
 203 ASKHEKILKSYGADVDFVHDAGVIEQIKSKYPNLOHVDAVGSSEDSIPEAYKVTADSL 262
 203 AGDEKATLFEKELG-----EYFIDFTKTDQMVAVEIATGVGA 240
 263 PATLLEVVPWTISIEPEIKDNVKIDITLLY-RASQELILGATR-FPASPEYHEATV 320
 241 HAVI-----NVSVESEAISTVSLTNGSTVVLVGLPRDACKSVFNGVVK 287
 321 FVKEFINLNGDIIHNNIKVFSNGLDVDPALTEGIEK 358
 288 SISIVGSYVGNRADTRALDFFSRGLVKAPIKILGLSE 325

RESULT 14

ADH2_PICST

ID ADH2_PICST STANDARD; PRT; 348 AA.

AC O13309; 000090;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE Alcohol dehydrogenase II (EC 1.1.1.1) (ADH 1).

OS ADH2 OR ADH1.
 OS Pichia stipitis (Yeast).


```

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Pichia.
OC NCBI_TaxID=4924;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 58785 / CBS 6054;
RA MEDLINE=98207839; PubMed=9546172;
RT Cho J.Y., Jeffries T.W.;
RT "Pichia stipitis genes for alcohol dehydrogenase with fermentative
RT and respiratory functions.";
RL Appl. Environ. Microbiol. 64:1350-1358(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CBS 5774;
RA MEDLINE=99019018; PubMed=9802210;
RA Passoth V., Schaefer B., Liebel B., Weierstall T., Kliner U.;
RA "Molecular cloning of alcohol dehydrogenase genes of the yeast Pichia
RA stipitis and identification of the fermentative ADH.";
RT Yeast 14:1311-1325(1998).
CC -1- FUNCTION: CONVERTS ETHANOL TO ACETALDEHYDE.
CC -1- CATALYTIC ACTIVITY: An alcohol + NAD(+) = an aldehyde or ketone +
CC NADH.
CC -1- COFACTOR: ZINC; BINDS 2 ZINC IONS, ONE ESSENTIAL FOR ACTIVITY,
CC THE OTHER FOR THE STRUCTURAL INTEGRITY.
CC -1- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, AF08244; AAC49990.1; -
DR EMBL, Y13238; CAAT3690.1; -
DR InterPro: IPR002328; ADH_zinc.
DR InterPro: IPR002085; Adh_zn_family.
DR Pfam: PF00107; adh_zinc; 1.
DR PROSITE: PS00059; ADH_ZINC; 1.
KW Oxidoreductase; Zinc; NAD; Multigene family.
FT METAL 44
FT METAL 44
FT METAL 67
FT METAL 98
FT METAL 101
FT METAL 104
FT METAL 112
FT METAL 154
FT METAL 154
FT METAL 41
FT METAL 41
SQ SEQUENCE 348 AA; 36565 MW; F6C813C98F56A148 CRC64;

Query Match 7.9%; Score 152.5; DB 1; Length 348;
Best Local Similarity 24.3%; Pred. No. 0.00049;
Matches 97; Conservative 51; Mismatches 136; Indels 115; Gaps 21;

QY 1 MSVPTQKAVITE--GDKAVVKTDSVPELKESTALVKEAVAGNPTD--WK-----48
DB 1 MSPTQKAVITEFTNGPLLYK-DIVPKPKRELLINVKYSGVCHTDLHAMKGDWPLDT 59
QY 49 HIAVKIGPBGSLICDIAGTVKLGPNASTDLKVPD-TGEGFVHG-----ASQTD 97
DB 60 KPLVGVGHEG-----AGVVVALGENV-TGMEIGYAGIKWINGSCLOCEYCVTAHESN 111
QY 98 -----PKNGAPAVYARVYRPLPYKSNLVTSTADEISEGPV-KNPSAASLPVSLTT 147
DB 112 CPNADISGYTHDSFOOYA-----TDAIOARIPIKGTDLALAPI--LC 154
QY 148 AGVSLCHLGSKN-----EMHPTPOHTPLLLINGCATAVGOOLIOVAKHINATYKIVTVA 203
DB 155 AGITVVKALKTAQLAQGW-----VAAGAGGAGLSLAIQYAKAMGRVVGIDGG 204

```

```

QY 204 SKKHEKLKSGVADGDPDHDGAVIEQIKSKYPNLOHVIDAGSDESDPEAKYKTADSLP 263
DB 205 ADKGE-PAKSLGA-EVF-----VDFLSSFD-----VADVLK 234
QY 264 ATLEVPWTIESEIRKDNVKIDITLLYRASGOEILLGATRPAA-----SPEYHEATV 319
DB 235 ATNGAHGVINVSVERAMQGVSD-----YVAPGTGVLVG---LPAGAKVASVSSVY 286
QY 320 KFKFNPPLNNGDIHNNIKVFSNGLDVDPALTEGICE 358
DB 287 RTIQKGSYVGNRADSAREIDFFETRLIKCPKIVGLSE 325

RESULT 15
ADH1_PICST STANDARD; PRT; 348 AA.
AC 000097;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Alcohol dehydrogenase I (Ec 1.1.1.1) (ADH 2).
GN ADH1 OR ADH2.
OS Pichia stipitis (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Pichia.
OC NCBI_TaxID=4924;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 58785 / CBS 6054;
RA MEDLINE=98207839; PubMed=9546172;
RT Cho J.Y., Jeffries T.W.;
RT "Pichia stipitis genes for alcohol dehydrogenase with fermentative
RT and respiratory functions.";
RL Appl. Environ. Microbiol. 64:1350-1358(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CBS 5774;
RA MEDLINE=99019018; PubMed=9802210;
RA Passoth V., Schaefer B., Liebel B., Weierstall T., Kliner U.;
RA "Molecular cloning of alcohol dehydrogenase genes of the yeast Pichia
RA stipitis and identification of the fermentative ADH.";
RT Yeast 14:1311-1325(1998).
CC -1- FUNCTION: CONVERTS ETHANOL TO ACETALDEHYDE AND PLAYS A MAJOR ROLE
CC IN XYLOSE FERMENTATION.
CC -1- CATALYTIC ACTIVITY: An alcohol + NAD(+) = an aldehyde or ketone +
CC NADH.
CC -1- COFACTOR: ZINC; BINDS 2 ZINC IONS, ONE ESSENTIAL FOR ACTIVITY,
CC THE OTHER FOR THE STRUCTURAL INTEGRITY.
CC -1- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, AF08245; AAC49991.1; -
DR EMBL, Y1397; CAAT3827.1; -
DR InterPro: IPR002328; ADH_zinc.
DR InterPro: IPR002085; Adh_zn_family.
DR Pfam: PF00107; adh_zinc; 1.
DR PROSITE: PS00059; ADH_ZINC; 1.
KW Oxidoreductase; Zinc; NAD; Multigene family.
FT METAL 44
FT METAL 44
FT METAL 67
FT METAL 98
FT METAL 101
FT METAL 101

```

FT METAL 104 104 ZINC (SECOND ATOM) (BY SIMILARITY).
 FT METAL 112 112 ZINC (SECOND ATOM) (BY SIMILARITY).
 FT METAL 154 154 ZINC (CATALYTIC) (BY SIMILARITY).
 SQ . SEQUENCE 348 AA; 36520 MM; 49C06B54D5350F4 CRC64;

Query Match 7.8%; Score 149.5; DB 1; Length 348;

Best Local Similarity 23.6%; Pred. No. 0.00082;

Matches 92; Conservative 51; Mismatches 150; Indels 97; Gaps 18;

```

QY 1 MSVPTTOKAVIIEGDKA-VVKTDVSVPELKEGTALVKEAVAGNPTD--WK-----H 49
Db 1 MSVPTTOKAVVFEENGGLPKDIPPTPKNEILLINKISGVCHTDLHANKGDPPLDTK 60
QY 50 IAVKIGPEGSLIGCDIAGTVVKLGPNASTDLKVG-DTGFVHGA-----SQTDPKNG 101
Db 61 LPLVGGHEG-----AGVVVGIGSNV-TGWELGDVAGIKMLNGSCINCERCFQHSDEPNC 112
QY 102 AFAEYARVYPPLEFKSNLTH-----STADEISEGPKVKNFESNASLP--VSLTTAGVSL 152
Db 113 AKADL-----SGYTHDGSFOOYATADAV-----QAARLPKGTDLQAAPIL 153
QY 153 CHHLSKMEWHPTPOHTPLLIWGGATVAGQOLIYAKHINAYTKITVASKKHEKLLK 212
Db 154 CAGITVYKALKTAQIQENWVCISGAGGGLSLAIQYAKAM-GPRVIALDGEKEGEFVK 212
QY 213 SYGADVDVYHDAGVIEQIKSKYPNLOHVIDAVGSEDSIPEAYKVTA DSLPATLLEVP 272
Db 213 SLGAEAYVDF-----TVSKDVIKDIQTAT-DGGPHALINV--- 246
QY 273 TIESIPEIRKDNKIDITLLYRASGOEILLGATRFPA SPP---YHEATYKPFVKFINPH 328
Db 247 ---SVSEKAIAQSCQ---YVRSTGTIVLVG--LPAGAKVAVPFDVAVKSTISIRGSY 295
QY 329 LNNGDIIHNNIKVFSNGILDVPALETGIIKE 358
Db 296 VGNRADSAEALDFTTRGLIKPIKVGIISE 325

```

Search completed: June 17, 2003, 10:49:14
 Job time : 19 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2003, 10:12:23 ; Search time 145 seconds

(without alignments)
525.775 Million cell updates/sec

Title: US-10-081-644-2

Perfect score: 1920
Sequence: 1 MSVPTQKAVIEGDKAVK.....ALTEGKKGKNNKVVAVRL 370

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Maximum Match 0%

Listing first 45 summaries

Database :

SPRENBL_21:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_ricent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacterioplasmid:*
17: sp_archaeoplasmid:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	309.5	16.1	348	3	Q9P7F4 schizosacch
2	277.5	14.5	339	10	Q94110 fragaria an
3	271	14.1	337	10	Q23939 fragaria ve
4	266.5	13.9	363	3	Q9Y7D0 aspergillus
5	264	13.8	313	16	Q8Y989 listeria mo
6	261	13.6	386	10	Q945P3 arabidopsis
7	260	13.5	313	16	Q92E39 listeria in
8	259	13.5	317	10	Q43677 vigina radia
9	248	12.9	309	10	Q94F67 helianthus
10	218	11.4	353	16	Q8UK00 streptomyce
11	214	11.1	313	16	Q85702 streptomyce
12	208	10.8	324	16	Q9KFE8 bacillus ha
13	200	10.4	309	17	Q8TSM3 mechanosarc
14	193	10.1	343	16	Q67374 aquifex aeo
15	189	9.8	308	16	Q98HL3 rhizobium 1
16	188	9.8	339	16	Q9L2A5 streptomyce

17	187.5	9.8	322	16	Q9KEB8 bacillus ha
18	187	9.7	359	16	Q8UW9 agrobacteri
19	182	9.5	329	10	Q9AYU1 triphysaria
20	176	9.2	329	16	Q9RUR7 streptomyce
21	176	9.2	331	17	Q97V32 sulfobolus
22	174.5	9.1	358	16	Q8UKD5 agrobacteri
23	173	9.0	328	16	Q9CHL2 lactococcus
24	172.5	9.0	339	16	Q9KEW1 bacillus ha
25	172	9.0	337	16	Q92R74 rhizobium m
26	171.5	8.9	338	16	Q8UGN9 agrobacteri
27	170.5	8.9	336	16	Q8YIY7 bruceella me
28	168	8.8	334	16	Q98J41 rhizobium 1
29	167.5	8.7	288	10	Q9SXP6 lithospermum
30	167	8.7	366	10	Q9LK96 arabidopsis
31	165	8.6	332	17	Q96YB5 sulfobolus
32	164.5	8.6	329	10	Q9SV68 arabidopsis
33	164.5	8.6	348	16	Q8UH60 agrobacteri
34	164	8.5	337	16	Q9KUG9 vibrio chol
35	163	8.5	335	5	Q8T122 dictyosteli
36	161.5	8.4	335	16	Q53146 mycobacteri
37	159.5	8.3	318	16	Q8Y2P5 arabidopsis
38	158.5	8.3	354	4	Q9BR44 homo sapien
39	158.5	8.3	363	16	Q985B6 rhizobium 1
40	158	8.2	326	16	Q98F10 rhizobium 1
41	157	8.2	328	5	Q45496 caenorhabdit
42	157	8.2	360	10	Q8RHH4 oryza sativ
43	156.5	8.2	8563	2	Q54297 streptomyce
44	156.5	8.2	10223	2	Q54296 streptomyce
45	156	8.1	328	16	P72043 mycobacteri

ALIGNMENTS

RESULT 1	Q9P7F4	PRELIMINARY;	PRT;	348 AA.
AC	Q9P7F4	01-OCT-2000 (TRENBLREL. 15, Created)		
DT	01-OCT-2000 (TRENBLREL. 15, Last sequence update)			
DT	01-OCT-2001 (TRENBLREL. 18, Last annotation update)			
DE	Putative dehydrogenase by similarity.			
GN	SPAC2B1P3.01.			
OS	Schizosaccharomyces pombe (Fission yeast).			
OC	Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;			
OC	Schizosaccharomycetales; Schizosaccharomycetaceae;			
OC	Schizosaccharomycetes.			
OX	NCBI_TaxID=4896;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=972H-;			
RA	Harris D., Wood V., Rajandream M.A., Barrell B.G.;			
RL	Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AL162531; CAB83005.1; -			
DR	InterPro: IPR02085; Adh_zn_family.			
Pf	Pfam: PF00107; adh_zinc1.			
SQ	SEQUENCE 348 AA; 37929 MW; 63B638A712AF1C2 CRC64;			
Query Match	16.1%; Score 309.5; DB 3; Length 348;			
Best local Similarity	28.2%; Pred. No. 7.3e-16;			
Matches	110; Conservative 49; Mismatches 160; Indels 71; Gaps 13;			
QY	8 KAVIEGDKAV-VKIDVSVPELKEGTALVKVEAVVGNPTDWMKHIYKIGPESIIACDIA 66			
DB	2 KAVIADGGVGVISDAPKTPPEKGEPLGRVIRVAFNPIDWTKLYNASIEKGTGVTGTFV 61			
QY	67 GTVVILGNASTDLKVGDTGFGFVAGASQTDPKNGAPAEVAVVPEPLFV--KSNLTHTA 124			
DB	62 AVEVAGGVGCVDRSKITGATVSGMAPG--PLDSNNAKRYITLDVNLVYFVKNITTPS-- 117			
QY	125 DEISGPNVNFESAASLPVSLTTAGVSLCHHLG-----SKMHPSTPQHTHP 172			
DB	118 -----QAVTLPLFTTASGLNQYGLPLPTDGSKNSAQQGM----- 155			

QY 173 LLIWGATAVGQOLIQVAKHINATKIVTASKEKELKSYGADVDFYHDAGVIEQIK 232
 ID 023939 PRELIMINARY; PRT; 339 AA.
 AC 023939;
 DT 01-JAN-1998 (Tremblrel. 05, last sequence update)
 DT 01-JUN-1998 (Tremblrel. 19, last annotation update)
 DE Ripening-induced protein (Fragment).
 OS Fragaria vesca (Woodland strawberry).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC Eurosid I; Rosales; Rosaceae; Rosoideae; Fragaria.
 NCBI_TaxID=57918;
 OX NCBI
 RN SEQUENCE FROM N.A.
 RC STRAIN=CV. REINE DES VALLEES;
 RA Nam Y.M., Tichit L., Marty I., Lelievre J.M.;
 RT "Isolation and characterization of cDNAs from genes differentially
 expressed during ripening of wild strawberry (Fragaria vesca L.).";
 RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ001445; CA04767.1;
 DR InterPro; IPR002085; Adh_zn_family.
 DR Pfam; PF00107; adh_zinc; 1.
 FT NON_TER 1
 SQ SEQUENCE 337 AA; 35914 MW; 4770A33FF83666FA CRC64;

RESULT 2

QY 173 LLIWGATAVGQOLIQVAKHINATKIVTASKEKELKSYGADVDFYHDAGVIEQIK 232
 ID 023939 PRELIMINARY; PRT; 339 AA.
 AC 023939;
 DT 01-DEC-2001 (Tremblrel. 19, last sequence update)
 DT 01-MAR-2002 (Tremblrel. 20, last annotation update)
 DE Putative quinone oxidoreductase (Fragment).
 OS Fragaria ananassa (Strawberry).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC Eurosid I; Rosales; Rosaceae; Rosoideae; Fragaria.
 NCBI_TaxID=3747;
 OX NCBI
 RN SEQUENCE FROM N.A.
 RC STRAIN=CV. CHANDLER;
 RA Lopez-Raez J.A., Redondo-Nevado J., Medina-Escobar N.,
 RT "A putative strawberry quinone oxidoreductase fruit ripening induced
 protein.";
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY048861; AAL06644.1;
 DR InterPro; IPR002085; Adh_zn_family.
 DR Pfam; PF00107; adh_zinc; 1.
 FT NON_TER 1
 SQ SEQUENCE 339 AA; 36194 MW; 475D2EEB3471640 CRC64;

Query Match 14.5%; Score 277.5; DB 10; Length 339;
 Best Local Similarity 29.7%; Pred. No. 2,2e-13;
 Matches 102; Conservative 52; Mismatches 115; Indels 75; Gaps 17;

2 SVPTTOKA-VIIEGDKA--VKTD--VSPELKEGTALVKEVAAGNPDMKH-IAY--- 52
 112 PLFYKSNLTHSTADEISEGPVKN--FESAASLPVSLTTA--GVALCHHLSKMEWHPS 169
 143 RV-----LAHKPKK-----LSFTEAASLPALAEHGLERALSAGK----- 180
 QY 170 THPLLIWGATAVGQOLIQVAKHINATKIVTASKEKELKSYGADVDFYHDAGVIE 229
 ID 023939 PRELIMINARY; PRT; 339 AA.
 AC 023939;
 DT 01-DEC-2001 (Tremblrel. 19, last sequence update)
 DT 01-MAR-2002 (Tremblrel. 20, last annotation update)
 DE Putative quinone oxidoreductase (Fragment).
 OS Fragaria ananassa (Strawberry).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC Eurosid I; Rosales; Rosaceae; Rosoideae; Fragaria.
 NCBI_TaxID=3747;
 OX NCBI
 RN SEQUENCE FROM N.A.
 RC STRAIN=CV. CHANDLER;
 RA Lopez-Raez J.A., Redondo-Nevado J., Medina-Escobar N.,
 RT "A putative strawberry quinone oxidoreductase fruit ripening induced
 protein.";
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY048861; AAL06644.1;
 DR InterPro; IPR002085; Adh_zn_family.
 DR Pfam; PF00107; adh_zinc; 1.
 FT NON_TER 1
 SQ SEQUENCE 339 AA; 36194 MW; 475D2EEB3471640 CRC64;

RESULT 3
 ID 023939 PRELIMINARY; PRT; 337 AA.
 AC 023939;
 DT 01-JAN-1998 (Tremblrel. 05, last sequence update)
 DT 01-JUN-1998 (Tremblrel. 19, last annotation update)
 DE Ripening-induced protein (Fragment).
 OS Fragaria vesca (Woodland strawberry).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC Eurosid I; Rosales; Rosaceae; Rosoideae; Fragaria.
 NCBI_TaxID=57918;
 OX NCBI
 RN SEQUENCE FROM N.A.
 RC STRAIN=CV. REINE DES VALLEES;
 RA Nam Y.M., Tichit L., Marty I., Lelievre J.M.;
 RT "Isolation and characterization of cDNAs from genes differentially
 expressed during ripening of wild strawberry (Fragaria vesca L.).";
 RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ001445; CA04767.1;
 DR InterPro; IPR002085; Adh_zn_family.
 DR Pfam; PF00107; adh_zinc; 1.
 FT NON_TER 1
 SQ SEQUENCE 337 AA; 35914 MW; 4770A33FF83666FA CRC64;

Query Match 14.1%; Score 271; DB 10; Length 337;
 Best Local Similarity 30.5%; Pred. No. 6,8e-13;
 Matches 105; Conservative 50; Mismatches 113; Indels 76; Gaps 19;

2 SVPTTOKA-VIIEGDKA--VKTD--VSPELKEGTALVKEVAAGNPDMKH-IAY--- 52
 23 SPSVKNKAWVSEYKGTSDVLKEDPSVAVPEIKEDQVLIKVAASINPVDFKRALGYFKD 82
 QY 53 KIEPESILGCDIAGTVKVGKPNASTDLKVGDTGFGVHGASQTD--KNGAFAYAYVP 111
 DB 83 TDSPLPTIPGYVAVGAVVAKGSGV--TKFKVGDVYDNLMTALVNTFRGSLAEY----- 136
 112 PLFYKSNLTHSTADE--ISEGPVKN--FESAASLPVSLTTA--GVALCHHLSKMEWHPS 165
 137 -----TADDEVLAHKP--KNLSFTEAASLPALAEHGLERALSAGK----- 179
 QY 166 TPQHTHPLLIWGATAVGQOLIQVAKHINATKIVTASKEKELKSYGADVDFYHDAGV 225
 DB 180 -----SVAVLGAGAGVGTHIIQLAKHVGASKVAATASKKDLRLTGADLAIDYTK 233
 QY 226 GVIEQIKSKVKNQHYIVAVGSEDSIPEAYK-----VTADSIPATLLEVP---MTIE 275
 DB 234 N-FEDLPEK--DVVYDAVGETDKAVKAVGSGKVVTVIGPATPPALIFVLTSKGSYLE 289
 QY 276 SIPEIRKDNVK--IDITILY-----RASGOEIL 302
 DB 290 KLRPYLESKGVKPVLDPTSPYPTKVAEAFGLSSRAIGKVVV 333

RESULT 4

QY 276 SIPEIRKDNVK--IDITILY-----RASGOEIL 302
 ID 023939 PRELIMINARY; PRT; 363 AA.
 AC 023939;
 DT 01-NOV-1999 (Tremblrel. 12, last sequence update)
 DT 01-NOV-1999 (Tremblrel. 12, last sequence update)
 DT 01-OCT-2001 (Tremblrel. 18, last annotation update)
 DE Enoyl reductase.
 OS Aspergillus terreus.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
 OC Eurotiiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
 OX NCBI
 RN SEQUENCE FROM N.A.
 RC STRAIN=ATCC0542;
 RA Kennedy J., Auclair K., Kendrew S.G., Park C., Vederas J.C.,

RA. Hutchinson C.R.;
 RT "Accessory Proteins Modulate Polyketide Synthase Activity During
 RT Lovastatin Biosynthesis.";
 RL Science 0:0-0(1999).
 DR EMBL; AF141925; AAD34554.1; -
 DR InterPro; IPR002085; Adh_zn_family.
 DR Pfam; PF00107; adh_zinc; 1.
 DR SEQUENCE 363 AA; 39511 MW; FDB9524DDB255713 CRC64;
 SO
 Query Match 13.9%; Score 266.5; DB 3; Length 363;
 Best Local Similarity 26.3%; Pred. No. 1,7e-12;
 Matches 103; Conservative 69; Mismatches 156; Indels 63; Gaps 17;
 QY 4 PTTQKAVII-EGKAVKTDVSVPELKEGTALVKAFAVAGNPTDMGHIAKIGPESSIG 62
 DB 9 PPGQTLTVNDHDEVTVWNAAPCPMLPRDOYTVKVAVAINSDTK-MGQFATPWAF 67
 QY 63 CDIACTVVKLGPNASTDLVKVDGTFGFGVAGASQDPRKNGAFAY---ARVYPLFKSN 118
 DB 68 TDVAGTVVAVGSDV-THIQGDRVYGAQNMCPRTDQAFSGYTVTRGRVA----- 119
 QY 119 LTHSTADEISEGVPKNPESASLPVSLITNAGVSLCHHLSKMEHPTST---PQHTP--L 173
 DB 120 -----KIPKG--LSFEPQAAALPAGISTAGLAM-KILGLPLP-SPSADQPPHSPRYV 168
 QY 174 LIMGATVAGQQLIOVAKHINATKIVTVASKKHEKLSKSYGADVDYHDAGVTEQIKS 233
 DB 169 LVYGSSTATATVTTQMQLR-LSGYTIPLAT-CSPHNPFLASRGAEVFDRAFNLAOTIT 226
 QY 234 -KYPNLQHVIVDAVGSSEDSIPEAYKVTADSLPATLLEVPMPTIESIPEE-IRKDNV 291
 DB 227 YTKNNLRVALDCITVNESTTFPCAATGRAAGHY-----VSLNPPEHATRMVTTDWT 280
 QY 292 LLYRASGQELIGATFP-----PASPETHAATVVKFKYINPHLNGDIHNNI 339
 DB 281 -----LGPITFGEGSTWPAFYRGSGSEERQFEDMDIRIAGOLVEDGLVHPL 329
 QY 340 KVFNSGLDDVPALTGEGKKNKVKYVRL 370
 DB 330 RVVQSGFDHIKQGMELVRKGEISGEKLVRL 360
 RESULT 5
 ID 08Y9B9 PRELIMINARY; PRT; 313 AA.
 AC 08Y9B9,
 01-MAR-2002 (TrEMBLrel. 20, Created)
 01-JUN-2002 (TrEMBLrel. 20, Last sequence update)
 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 GN Hypothetical protein lmo0613.
 OS Listeria monocytogenes.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
 OC Listeriaceae; Listeria.
 NCBI_TaxId=1639;
 RX MEDLINE=21537279; PubMed=11679669;
 RA Baquero F., Frangoul L., Buchrieser C., Rusniok C., Amend A.,
 RA Glaeser P., Berche P., Bioecker H., Brandt P., Chakraborty T.,
 RA Charbit A., Chetoui F., Couve E., de Daruvar A., Deboux P.,
 RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Duesenget O.,
 RA Entian K.-D., Fath H., Garcia-del Portillo F., Garrido P.,
 RA Gautier L., Gobel W., Gomez-Lopez N., Hain T., Haut J., Jackson D.,
 RA Jones L.-M., Kaerst U., Kretz J., Kuhn M., Kunst F., Kurapkak G.,
 RA Maduno E., Maitounan A., Mata Vicente J., Ng E., Nedjari H.,
 RA Nordiek G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,
 RA Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
 RA Vazquez-Boland J.-A., Voets H., Wehland J., Cossart P.;
 RT "Comparative genomics of Listeria species.";
 RL Science 294:849-852(2001).
 DR EMBL; AL591976; CAC98691.1; -

DR Listeria; LMO0613; -
 DR InterPro; IPR002085; Adh_zn_family.
 DR Pfam; PF00107; adh_zinc; 1.
 KW Hypothetical protein; Complete proteome.
 DR SEQUENCE 313 AA; 33940 MW; 0F1012C59A23621D CRC64;
 SO
 Query Match 13.8%; Score 264; DB 16; Length 313;
 Best Local Similarity 29.0%; Pred. No. 2,1e-12;
 Matches 102; Conservative 56; Mismatches 120; Indels 74; Gaps 18;
 QY 8 KAVIIE--GDKAVKTDVSVPELKEGTALVKAFAVAGNPTDMK---HIAKIGPE-GS 59
 DB 2 KAVIENYGGKEELKEKEVAMPKAGNIVYKAAATISINPDKLREGVLKQMDWEFPI 61
 QY 60 IIGCIIAGTVVKLGPNASTDLVKVDGTFGFGVAGASQDPRKNGAFAYAVYPLFKSNL 119
 DB 62 IIGMDVAVGISSEVGGV-TDWMKVGSEVF-----APETTRFGTYAETVAV-----DDHL 109
 QY 120 THSTADEISEGVPKNPESASLPVSLITNAGVSLCHHLSKMEHPTSTPQHTPLIHWGA 179
 DB 110 LAPLEGIS-----FDEAASIPLAGITMQLPDIH-AKL-----QKGEKVLHAGA 154
 QY 180 TAVGQQLIOVAKHINATKIVTVASKKHEKLSKSYGADVDYHDAGVIEQIKSKYPNLQ 239
 DB 155 GGVGTIAQLAHAGA--EVITTAQKNEHLKSGADQVIDYKEV---NFKVDVLSID 208
 QY 240 HYTDVAGSSEDSIPEAYKVTADSLPATLLEVPMPTIESIPEERKDNVKTIDITLY-RAG 298
 DB 209 VVFDTWGGQIE-TDSYDVJKEG-TGLRISIV-----GLSNEBRAKENVATATGIMQPN 261
 QY 299 OEI-----LIGATFP-----ASPETHAATVVKFKIN 326
 DB 262 EQLKELGLANKTIKPIYAT-FPPEKGVDAHALSTHVAQKIVISFN 312
 RESULT 6
 ID 0945P3 PRELIMINARY; PRT; 386 AA.
 AC 0945P3,
 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE At1g23740/F508.27.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 NCBI_TaxId=3702;
 RX [1]
 RP SEQUENCE FROM N.A.
 RA Shin P., Chen H., Cheuk R., Kim C.J., Koesema E., Meyers M.C.,
 RA Banh J., Bowser L., Carinici P., Dale J.M., Goldsmith A.D.,
 RA Hayashizaki Y., Ishida J., Jiang P.X., Jones T., Kamiya A.,
 RA Karlin-Neumann G., Kawai J., Lam B., Lee J.M., Lin J., Liu S.X.,
 RA Miranda M., Narusaka M., Nguyen M., Onodera C.S., Palm C.J.,
 RA Pham P.K., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,
 RA Tang C.C., Toriumi M., Yamada K., Yamamura Y., Yu G., Yu S.,
 RA Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
 RT "Arabidopsis cDNA clones.";
 RL Submitted (AUG-2001) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Cheuk R., Chen H., Kim C.J., Meyers M.C., Shin P., Banh J.,
 RA Bowser L., Carinici P., Chang E., Dale J.M., Goldsmith A.D.,
 RA Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,
 RA Kawai J., Lam B., Lee J.M., Lin J., Miranda M., Narusaka M.,
 RA Nguyen M., Onodera C.S., Palm C.J., Quach H.L., Sakurai T., Satou M.,
 RA Seki M., Southwick A., Tang C.C., Toriumi M., Wu H.C., Yamada K.,
 RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
 RA Ecker J.R.;
 RT "Arabidopsis ORF clones.";
 RL Submitted (APR-2002) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AF411799; AAL06488.1; -

QY 286 VKIDITLVRASGQELTGATRPSPPEYHEATVKEVKFINPHLNNGDI 334
 DB 266 DLUEKRLPYLNGQ---VKPILDPKSPFPSPQTEVAEFTLKTNRATGKV 311

RESULT 9

Q94F67 PRELIMINARY; PRT; 309 AA.

AC 094F67; 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)

DE Guanine oxidoreductase-like protein.
 GN HAPLOOP.
 OS Helianthus annuus (Common sunflower).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Asteridae; euasterids II; Asterales; Asteraceae; Asteroideae;
 Heliantheae; Helianthus.
 NCBI_TaxID=4232;

RP SEQUENCE FROM N.A.
 RA Vugrek O.;
 RT "Assaying the quality of size-selected root-specific sunflower cDNA
 libraries";
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF384244; AAK66565.1; -
 DR InterPro; IPR002085; Ach_zn_family.
 DR Pfam; PF00107; ach_zinc; 1.
 SQ SEQUENCE 309 AA; 33190 MW; EA033406CCE6D83 CRC64;

Query Match 12.9%; Score 248; DB 10; Length 309;
 Best Local Similarity 31.2%; Pred. No. 3, 6e-11;
 Matches 82; Conservative 34; Mismatches 101; Indels 46; Gaps 10;

QY 15 DRAVVTDSVPELKEGTALVKEAAGNPTDKHIAVKIG-----PEGSTLGGDING 67
 DB 13 DVLKATDVAPEIKDDQVLVKAANAANPVYKR---RLGYKALDSPLPIIRGDDVG 69
 QY 68 TVVKLGPNASTDLKVDGTFGFGVHGASQTDPKN-GAFEAARYPPPL-FYKSNLTHSTA 124
 DB 70 VVLKVGSGYK-DLKEGDDEVYGDINDKLEGPQFGTLAEYTAVERLALKPKNL----- 123
 QY 125 DELSEPVKPEASALPVSITTA--GVSILCHLGSKMENHSTPQHTPBLIMGATAY 182
 DB 124 -----DFIOAALPLAIEFYAYEGLERAKPESEK-----TIIVINGAGGV 162
 QY 183 GQQLIOVAKHINAYTKIVTASCKHEKLSYGADDPFDYHDAGVTEQIKSKYENLOHYI 242
 DB 163 GSFIIOLAKHVGASKVAATSTGKLEKLSLADVAIDITKEN-FEDLPDKT---DVY 218
 QY 243 DAVGSEDSIPEAYKVTADSLPAT 265
 DB 219 DAIGPEKALKAVNETGVANVSIT 241

RESULT 10

Q8UK00 PRELIMINARY; PRT; 353 AA.

AC 08UK00; 01-JUN-2002 (TREMBlrel. 21, Created)
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Zinc-binding oxidoreductase.

GN ATU5324 OR AGR PAT 466.
 OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
 OS Plasmid AT.

OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Rhizobium.
 OC NCBI_TaxID=176299;
 RN (1)
 RP SEQUENCE FROM N.A.

RX MEDLINE=21608550; PubMed=11743193;

RA Wood D.W., Serubaj J.C., Kaul R., Monks D.E., Kitajima J.P.,
 RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F., Jr., Woo L.,
 RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,
 RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
 RA Kutayavin T., Levy R., Li M.-J., McClelland E., Palmeri A.,
 RA Raymond C., Rouse G., Seneffmachak C., Wu Z., Romero P.,
 RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
 RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
 RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
 RA Nester E.W.;
 RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
 C58";
 RL Science 294:2317-2323 (2001).

RN (12)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21608551; PubMed=11743194;
 RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
 RA Qurollo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
 RA Hounmel K., Gordon J., Vaudin M., Iarchouk O., Epp A., Liu F.,
 RA Mollam C., Allinger M., Dougherty D., Scott C., Lappas C., Markelz B.,
 RA Flanagan C., Crowell C., Gursan J., Lomo C., Sear C., Strub G.,
 RA Cielo C., Slater S.;
 RT "Genome sequence of the plant pathogen and biotechnology agent
 RT Agrobacterium tumefaciens C58";
 RL Science 294:2323-2328 (2001).
 DR EMBL; AE008954; AAL46012.1; -
 DR EMBL; AE007901; AAK90698.1; -
 DR Plasmid; Complete proteome.
 SQ SEQUENCE 353 AA; 36922 MW; 618FBE7361FPA1A5 CRC64;

Query Match 11.4%; Score 218; DB 16; Length 353;
 Best Local Similarity 29.9%; Pred. No. 9, 6e-09;
 Matches 80; Conservative 36; Mismatches 106; Indels 46; Gaps 10;

QY 8 KAVIIEGDAVAVKTVSPEL-----KEGTALVKEAAGNPTD-----WKHI 50
 DB 13 EAMMEIMKAVAVLHEHGESEVLSYERAPRPAASGGLVLRVHAAGNPFDDVLRDQRYTL 72
 QY 51 AVKIGPEG---ILGCDIAGTVKLGPNASTDLKVDGTFGFGVHGASQTDPKNKAFAEA 107
 DB 73 PPEMOPEPTPLILGDSGVVAALGDGVA--FSVGDDEVFAWVRPEPDLMOGSGAYAEYV 131
 QY 108 RYVPLFYKSNLTHSTADISSEGPVK-NFESAASLPVSLTTAG---VSLCHLGSME-- 161
 DB 132 TV-----PASELAKPAGCIDIDIOAGAWMSLITMQLFVLDIGHAPNFGQSF 178
 QY 162 WHPSTFQHTPPLIMGATAVGQQLIOVAKHINAYTKIVTASCKHEKLSYGADDPVD 221
 DB 179 RHPSPILQKTVLVNAGGSGVGHIALQLAKMKGAR--VIAVAGSGHEALLRLAGDQIID 236
 QY 222 YHDAGVIEQIKSKYENLOHYIDAVGSED 249
 DB 237 YTKTAA---ETAEDVDLVIDAVGGSN 260

RESULT 11

O85702 PRELIMINARY; PRT; 313 AA.

AC 085702; 01-NOV-1998 (TREMBlrel. 08, Created)
 DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Putative oxidoreductase (Putative zinc-binding oxidoreductase).

GN SC00780 OR 3SCPE60.12.
 OS Streptomyces lividans, and
 OS Streptomyces coelicolor.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomycetes.

OC NCBI_TaxID=1916, 1902;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC SPECIES=5.lividans; STRAIN=1326, AND AJ100;

RA Altenbuchner J.;
 RT "Amplifiable element AUD4 from *Streptomyces lividans* 66.";
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC SPECIES=S.coelicolor; STRAIN=A3(2);
 RA Seeger K.J., Harris D.;
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC SPECIES=S.coelicolor; STRAIN=A3(2);
 RA Cardeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RC SPECIES=S.coelicolor; STRAIN=A3(2);
 RX MEDLINE=97000351; PubMed=8643436;
 RA Redendach M., Kieser H.M., Denapate D., Eichner A., Cullum J.,
 RA Kinashi H., Hopwood D.A.;
 RA "A set of ordered cosmids and a detailed genetic and physical map for
 the 8 Mb *Streptomyces coelicolor* A3(2) chromosome.";
 M1. Microbiol. 21:77-96(1996).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC SPECIES=S.coelicolor; STRAIN=A3(2) / M145;
 RA Bentley S.D., Chater K.F., Cardeno-Tarraga A.M., Challis G.L.,
 RA Thomson N.R., James K.D., Harris D.B., Quail M.A., Kieser H.,
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornby T., Howarth S.,
 RA Huang C.H., Kieser I., Larke L., Murphy L., Oliver K., O'Neill S.,
 RA Rabbittowitch E., Rajandream M.A., Rutherford K., Rutter S.,
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
 RA Warren T., Wietorrek A., Woodward J., Barrell B.G., Parkhill J.,
 RA Hopwood D.A.;
 RT "Complete genome sequence of the model actinomycete *Streptomyces*
 coelicolor A3(2).";
 RL Nature 417:141-147(2002).
 DL EMBL; AF072709; AAC25771.1;
 DL EMBL; AL445945; CAC14345.1;
 DR HSSP; P28304; 1QOR.
 DR InterPro: IPR002085; Adh zn family.
 DR InterPro: IPR002364; QOR_zeta_crystal.
 DR Pfam; PF00107; adh_zinc; 1.
 DR PROSITE; PS01162; QOR_ZETA_CRYSTAL; 1.
 DR SEQUENCE 313 AA; 33213 MW; 4482871228BB4A84 CRC64;
 SQ

Query Match 11.1%; Score 214; DB 16; Length 313;
 Best Local Similarity 26.5%; Pred. No. 1.6e-08;
 Matches 86; Conservative 53; Mismatches 122; Indels 64; Gaps 17;

QY 6 TORAV--IIEGKAVVKTIVSVPELKEGTALVYKVAAGNPTDMKHIAYK--IGPEGSI 60
 DB 3 TMAIGDVLGPEVLEKEVPEREPNEVLVVRAGVNPDMKRAATGGLGEPFV 62
 QY 61 LGCDIAGTVYKLGPNASTDLKVGDTG---FVHGASQDTPKNGAPAEIARYPPLEFY 115
 DB 63 LGMDVSGVSVGVA--FRPGDEVEGMLSYPPGH-----GSHAEVVAAPARTFT 112
 QY 116 K--SNLHSTADEISEGPNVKNFESAASLPVSLTTAGVSLCHLGSKMEHPSPTQHNP 173
 DB 113 RKPSGIH-----VOAGALPLVSLTAMQALV---ERAEVQPG-----QRV 149
 QY 174 LIMGATAVGQQLIQVAKHINAYTKITVVAASKHEKLKSYGADVDVFDYHAGVIEQIS 233
 DB 150 LVHAAAGVGVAQAQAKAGARV-IGTASAAKHE-FLRSGADETVDYRETDPAAVVK- 206
 QY 234 KYPLQGVIVAVGSEDSIPAEYKTAADSLPATILEVYPMITESIPER-----IRKDNKI 288
 DB 207 ---DADVLDITIGDITSL-RSLRVLKRG--GVVVSILPVGSDDEYBEADRLGVAVVMLV 260
 QY 289 DITLLYRASGOEI--LIGATRFPAS 311
 DB 261 DAD---RAGMEELRLVAGAGLRAT 282

RESULT 12
 OQKFV8
 ID OQKFV8 PRELIMINARY; PRT; 324 AA.
 AC OQKFV8;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, last annotation update)
 DE Quinone oxidoreductase.
 GN BH0363.
 OS Bacillus halodurans.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
 OC Bacillaceae; Bacillus.
 OX NCBI_TaxId=86655;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C-125 / JCM 9153;
 RX MEDLINE=20512582; PubMed=11058132;
 RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
 RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
 RA Horikoshi K.;
 RT "Complete genome sequence of the alkaliphilic bacterium *Bacillus*
 halodurans and genomic sequence comparison with *Bacillus subtilis*.";
 RL Nucleic Acids Res. 28:4317-4331(2000).
 DL EMBL; AP001508; BAB04082.1;
 DR HSSP; P28304; 1QOR.
 DR InterPro: IPR002085; Adh zn family.
 DR InterPro: IPR002364; QOR_zeta_crystal.
 DR Pfam; PF00107; adh_zinc; 1.
 DR PROSITE; PS01162; QOR_ZETA_CRYSTAL; 1.
 DR Complete proteome.
 SQ SEQUENCE 324 AA; 34754 MW; 96690EC3445B79C5 CRC64;
 SQ

Query Match 10.8%; Score 208; DB 16; Length 324;
 Best Local Similarity 26.4%; Pred. No. 5e-08;
 Matches 90; Conservative 53; Mismatches 132; Indels 66; Gaps 15;

QY 8 KAVIIE--GDKAVVKTIVSVPELKEGTALVYKVAAGNPTDMKHIAYKIGPEGSI-- 61
 DB 2 KAVIVAFGPEHMMEDVSTPIKENEVLIKVKNSVNFAD--IKARYGRKGTLPFI 58
 QY 62 -GCDIAGTVYKLGPNASTDLKVGDTGFGVHGASQDTPKNGAPAEIARYPPLEFYSLNT 120
 DB 59 PGIDAAGYVEKRGREVS-DIQVGQRIAR-----PKSGSYAEV-----VVAESLV 103
 QY 121 HSTADEISEGPNVKNFESAASLPVSLTTAGVSLCHLGSKMEHPSPTQHNP 180
 DB 104 FPIPDEI-----NFKTAAASPI-----VSFLSH--RLIVVAGMERGESVLTVAAG 148
 QY 181 AVGQQLIQVAKHINAYTKITVVAASKHEKLKSYGADVDVFDYHAGVIEQISKYPNLQ 240
 DB 149 GVGTTAIQVAKLIGAGTIVGVSADKITAAGSGADDEVYCYEEDFTKSV-----OE 201
 QY 241 VIDAVSEDSIPAEYKTAADSLPATILEVYPMITESIPERIRKDNKIDKID--TLVYA 296
 DB 202 MTNGGVVD-----IILDVSGVSTKSLQCLARYGLVHFGNSSGAGITIKITIDLHA 253
 QY 297 SGOEIL--LIGATRFPASPEYHATVFKFIFINPHLNGDI 334
 DB 254 SCRSVIGFSLGTR-KCKEHLKETAHYLV---PYLASKL 290

RESULT 13
 OQTSM3
 ID OQTSM3 PRELIMINARY; PRT; 309 AA.
 AC OQTSM3;
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)
 DT 01-JUN-2002 (TrEMBLrel. 21, last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, last annotation update)
 DE NADPH:quinone reductase.
 GN MA0773.
 OS Methanobacterium acetivorans.

OC Archaea; Euryarchaeota; Methanococci; Methanosarcinales;
 OC Methanosarcinaceae; Methanosarcina.
 NCBI_TaxID=2214;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN=CZA / ATCC 35395 / DSM 2834;
 RX MEDLINE=21929760; PubMed=11932238;
 RA Galagan J.E., Nusbbaum C., Roy A., Endrizzi M.G., Macdonald P.,
 RA Fitzhugh W., Calvo S., Engels R., Smirnov S., Atchaeo D., Brown A.,
 RA Allen N., Naylor J., Strange-Thomann N., DeRellano K., Johnson R.,
 RA Linton L., McEwan P., McKernan K., Talamas J., Tirrell A., Ye W.,
 RA Zimmer A., Barber R.D., Cann I., Graham D.E., Graham D.A., Guss A.M.,
 RA Hedderich R., Ingram-Smith C., Kuetner H.C., Krzycki J.A.,
 RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
 RA Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,
 RA Ferry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.,
 RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
 RA Metcalf W.W., Birren B.;
 "The genome of Methanosarcina acetivorans reveals extensive metabolic
 and physiological diversity";
 Genome Res. 12:532-542(2002).
 DR EMBL; AE010739; AM04212.1; -.
 KW Complete proteome.
 SQ SEQUENCE 309 AA; 33158 MW; D56D00207DA241D CRC64;
 Query Match 10.4%; Score 200; DB 17; Length 309;
 Best Local Similarity 26.2%; Pred. No. 2e-07;
 Matches 79; Conservative 52; Mismatches 120; Indels 50; Gaps 15;
 QY 8 KAVIIE--GDKAVKTDVSVPELKEGTALVKEAVAG--NPTDMKIAVKG--PREGSL 61
 DB 2 KAIRIHFGGKEVWKRE-DIPQPOGTGEIRIRIIAGVVPMDKIRSGMIGMPLMTM 60
 QY 62 GCCTAGTVVVKLGPNASTDLKVGDTGFGFVAGASQTPKNAFAYRVPPLPKSLTH 121
 DB 61 GIDVAGVVAQGP-GEVSPQGEVVF-----AKVSGGSVAYETV----- 101
 QY 122 STADESEGPVK-NFESAASLPVSLTTAGVSLCHHGSKMEKMPSPQTHPLLLINGCAT 180
 DB 102 -NSAQRKPKSGFESBAIPYAGLAWSLFDIAGL-----EKQSVLHGAAG 151
 QY 181 AVGGQILQVAKHINAYTKITVASKKHKLKSGYGDVDVYDAGVIEQIKSKYPNLQH 240
 DB 152 GVGSFAVQFARMKGA--VIGTASEKMKQFLKISGIDEFDYK---QGFEDAAGKVDV 205
 QY 241 VIDAVSEDSIPAYKYTADSLPATLLEVPMPTIESIPERIKONKIDITLLYRSGOE 300
 DB 206 VLDITGG-DTFDRSMGLKPG--GLVSTVARIPEGVPE---KXGVHAQ-TLMTRADGEE 258
 QY 301 I 301
 DB 259 I 259
 RESULT 14
 067374 PRELIMINARY; PRT; 343 AA.
 AC 067374;
 DT 01-AUG-1998 (Tremblrel. 07, Created)
 DT 01-AUG-1998 (Tremblrel. 07, Last sequence update)
 DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
 DE Alcohol dehydrogenase.
 GN ADHI OR AQ.1362.
 OS Aquifex aeolicus.
 OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.
 NCBI_TaxID=63363;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN=VFS;
 RX MEDLINE=98196666; PubMed=9537320;
 RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
 RA Graham D.E., Overbeek R., Sneed M.A., Keller M., AuJay M., Huber R.,
 RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;

RT "The complete genome of the hyperthermophilic bacterium Aquifex
 RT aeolicus";
 RL Nature 392:353-358(1998).
 DR EMBL; AE000736; AAC07327.1; -.
 DR HSSP; P28304; IQOR.
 DR InterPro; IPR002085; Adh_zn_family.
 DR Pfam; PF00107; adh_zinc; 1.
 KW Complete proteome.
 SQ SEQUENCE 343 AA; 38043 MW; EA77D5B7D0B5F8B CRC64;
 Query Match 10.1%; Score 193; DB 16; Length 343;
 Best Local Similarity 28.7%; Pred. No. 8e-07;
 Matches 77; Conservative 43; Mismatches 94; Indels 54; Gaps 14;
 QY 8 KAVIIEG----DKAVKTDVSVPELKEGTALVKEAVAGNPNP-WGHI-AVKIGPE-GSI 60
 DB 2 KAVIITGFGGIEELKVEEPFPEPEKDEVLIRKAVANLHLDIWRGALPVKPELPHI 61
 QY 61 LGCDIAGTVVVKLGPNASTDLKVG-----DTGFGFVAGASQTD----PKNG 101
 DB 62 LGSVSGVVERKVG-SLVKNVKEGEVITAPGLSCGVCWDCGGRNHCXDYDILGLKNG 120
 QY 102 APAEYAVVPPFLPKSLNTHSTADEISEGPVK--FESAASLPVSLTTAGVSLCHHLSK 159
 DB 121 GYAEYVAV-----PARNVKKP-KNLSPFEASVPLFLT---VMNALVDK 162
 QY 160 MEMHPESTPQTHPLILMGATAGVQQLIQAACHINAYTKITVASKKHKLKSYADAV 219
 DB 163 AQIKRPS-----RVFLWAGSSGVAGIQLAARFNAFV-ITTAGNEKAKKCKELGADLV 216
 QY 220 FVYHDAGVIEQIKSKY-PNLQHYIDAVG 246
 DB 217 FNVHKDQVVKRVEVKEGVVDVVDHIG 244
 RESULT 15
 098HL3 PRELIMINARY; PRT; 308 AA.
 AC 098HL3;
 DT 01-OCT-2001 (Tremblrel. 18, Created)
 DT 01-OCT-2001 (Tremblrel. 18, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Probable oxidoreductase.
 GN MR2819.
 OS Rhizobium loti (Mesorhizobium loti).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Phyllobacteriaceae; Mesorhizobium.
 NCBI_TaxID=381;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN=MAFP303099;
 RX MEDLINE=21082930; PubMed=11214968;
 RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
 RA Watanabe A., Idegawa K., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
 RA Mochizuki Y., Nakayama S., Nakazaki N., Shimo S., Sugimoto M.,
 RA Takeuchi C., Yamada M., Tabata S.;
 "Complete genome structure of the nitrogen-fixing symbiotic bacterium
 Mesorhizobium loti";
 RL DNA Res. 7:331-338(2000).
 DR EMBL; AP003000; BAB49853.1; -.
 DR InterPro; IPR002085; Adh_zn_family.
 DR InterPro; IPR002364; QOR_zeta_crystal.
 DR Pfam; PF00107; adh_zinc; 1.
 DR PROSITE; PS0162; QOR_ZETA_CRYSTAL; 1.
 KW Complete proteome.
 SQ SEQUENCE 308 AA; 32146 MW; 4E56DFDB90635808 CRC64;
 Query Match 9.8%; Score 189; DB 16; Length 308;
 Best Local Similarity 29.0%; Pred. No. 1.4e-06;
 Matches 72; Conservative 28; Mismatches 100; Indels 48; Gaps 10;
 QY 16 KAVKTDVSVPEL-----KEGTALVKEAVAGNPTDMK-HIAY--KIGPEGSL 61

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 06:23:31 ; Search time 3073 Seconds

(without alignments)
10540.652 Million cell updates/sec

Title: US-10-081-644-1

Perfect score: 1113
Sequence: 1 atgcagctccacacaccca.....agtatgtccagctataa 1113

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

1 number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: gb_ba:*
2: gb_hlg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pac:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_scs:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pac:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_scs:*
28: em_un:*
29: em_vl:*
30: em_hlg_hum:*
31: em_hlg_inv:*
32: em_hlg_ocher:*
33: em_hlg_mus:*
34: em_hlg_pln:*
35: em_hlg_rtd:*
36: em_hlg_mam:*
37: em_hlg_vrt:*
38: em_sy:*
39: em_higo_hum:*
40: em_higo_mus:*
41: em_higo_ocher:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	359	32.3	29634	8	YSC9122
2	357.8	32.1	1944	8	SCN113AC
3	357.8	32.1	43481	8	SCN113AC
4	347.8	31.2	316613	8	SCN113AC
5	252.4	22.7	1083	11	CNS06FEH
6	103.6	9.3	942	11	CNS06KAB
7	103.2	9.3	5355	8	YSC9122
8	53.2	4.8	1506	8	CCTORD
9	48.2	4.3	109375	8	AC003672
10	47.6	4.3	98734	2	PFMAL1P2
11	47.4	4.3	13844	1	AE007561
12	47.4	4.3	92357	9	HS1106901
13	47.4	4.3	119082	2	AC008057
14	47.4	4.3	169494	2	AC007630
15	47.4	4.2	84472	2	AC096684
16	47.4	4.2	179553	2	AC024253
17	46.4	4.2	52514	2	AC116551
18	46.2	4.2	5537	3	AP465308
19	45.4	4.1	9553	5	AF134321
20	44.6	4.0	11327	1	AE006787
21	44.6	4.0	163243	9	AL513304
22	44.2	4.0	71553	2	AC096997
23	44.2	4.0	186935	2	AC022322
24	44.2	4.0	175001	2	AC126846
25	43.8	3.9	157695	2	AL136442
26	43.8	3.9	176480	2	AC104805
27	43.8	3.9	325350	1	AP004823
28	43.6	3.9	264522	2	AC090437
29	43.2	3.9	183413	2	AC131200
30	43.2	3.9	215411	2	AC122304
31	42.8	3.8	108476	2	AC094507
32	42.8	3.8	116696	3	PFMAL1P3
33	42.6	3.8	4111	10	MM0278580
34	42.6	3.8	125026	10	AC091616
35	42.6	3.8	178273	2	AC005308
36	42.6	3.8	179780	2	AC127126
37	42.6	3.8	183419	2	AC116417
38	42.6	3.8	241870	2	AC122248
39	42.6	3.8	269050	1	AP000987
40	42.6	3.8	276184	2	AC127414
41	42.6	3.8	321003	2	PFMAL1P3
42	42.2	3.8	53150	2	AC116984
43	42.2	3.8	170300	2	AC116888
44	42.2	3.8	301050	1	AP003130
45	42.2	3.8	343590	1	AP003359

ALIGNMENTS

RESULT 1
LOCUS YSC9122 29634 bp DNA linear PLN 01-AUG-1997
DEFINITION Saccharomyces cerevisiae chromosome XII cosmid 9122.
ACCESSION U22383 Y13138
VERSION U22383.1 GI:2264349
KEYWORDS
SOURCE
ORGANISM Saccharomyces cerevisiae.
Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE
AUTHORS Johnston, M., Hillier, L., Riles, L., Albermann, K., Andre, B.,
Ansong, W., Benes, V., Bruckner, M., Delius, H., Dubois, B.,
Dusterhoft, A., Entian, K.D., Floeth, M., Goffeau, A., Heblung, U.,

Pred. No. is the number of results predicted by chance to have a

TITLE
 JOURNAL
 MEDLINE
 PUBMED
 9169871
 2 (bases 1 to 29634)
 Kisten, J.
 The sequence of *S. cerevisiae* cosmid 9122
 Unpublished (1994)
 3 (bases 1 to 29634)
 Waterston, R.
 Direct Submission
 Submitted (08-MAR-1995)
 4 (bases 1 to 29634)
 Cherry, J.M.
 Direct Submission
 Submitted (18-JUL-1997) Saccharomyces Genome Database
 5 (bases 1 to 29634)
 Cherry, J.M.
 Direct Submission
 Submitted (28-JUL-1997) Saccharomyces Genome Database
 On Jul 16, 1997 this sequence version replaced gi:717070.
 COMMENT
 Sequenced by:
 Genome Sequencing Center
 Department of Genetics, Washington University,
 St. Louis, MO 63110, USA
 e-mail: yeast-curator@genome.stanford.edu
 Curated by:
 Saccharomyces Genome Database
 URL: http://genome-www.stanford.edu/
 e-mail: yeast-curator@genome.stanford.edu

FEATURES
 source
 1. 29634
 /organism="Saccharomyces cerevisiae"
 /strain="S288C (AB972)"
 /db_xref="taxon:4932"
 /map="XII"
 51..7937
 /gene="YLR454W"
 51..7937
 /gene="YLR454W"
 /codon_start=1
 /evidence=not_experimental
 /product="Ylr454wp"
 /protein_id="AAB64718.1"
 /db_xref="GI:717075"
 /translation="MMFPIINVLLYKMLFVAVPLMSCKTLRLGINTWLNFLKE
 IGLSLSDGTVLKSVFAPVPERKLFKGLRIDSCKSSNDLAKELPREESTFETEP
 EDNGGFSIKLISLSQYWLNGVTIILEDQVANNDDTIKRFEPFLSINDSKITLRF
 DSFLRLKLNNGOTTIADALFIVNTNLIGBINPLKDGQVGLDKLGLNLPNNILN
 LFLKLNVDLMSNEKRLQRLADTKNEBKEDLAKMDLVYAMEKEDVRIKPLKE
 MNTVTVKIQIDPLTNHBLGMMKYISYVLSVINENFRFRMEKGYTLIFEEER
 DSPEFSIIMARENIYLNLRKSHAKOLKTIIEIPVSI FGETNLFSQKPLSNHLK
 AKLENAITINIGNISLTITDMDPVNISFKELSNIKYFTSSCPNKTKLKNSHYK
 LTRRVLDFPKFCFLINMKTLDLPKVIINDKDLIIGKSVFISHSKRYTLGN
 NMBEKEETQHLFYESHNVVELDMQLQHLIKQKQKHTLKVDSIAIEKVQLPDI
 LGSANADITLMDLSELPTMVLSELVNLSQLANVENEFKEFEYFAANLQMK
 ACSNNAKCLQKQKEILPSDFMFOQLPDFDIYIKINRISITGLASVMPDPVSSV
 DSOSSKDLIDGKLRKYNCTVEKQLALFDGKQMNKIGSSNATWRSGLNFSKDN
 KQNPNNKSSIALDDISTSDATEVNLNMINLVNDITTSITGETPEVSEBELSTKTVS

Neighboring Sequence:
 This sequence includes nucleotides 1-22200 of cosmid 9122. This is
 the rightmost cosmid of chromosome 12. The cosmid on the left is
 9324. This sequence overlaps the right telomeric sequence,
 accession number Z73377, starting at nucleotide 28299.
 Location/Qualifiers
 1. 29634
 /organism="Saccharomyces cerevisiae"
 /strain="S288C (AB972)"
 /db_xref="taxon:4932"
 /map="XII"
 51..7937
 /gene="YLR454W"
 51..7937
 /gene="YLR454W"
 /codon_start=1
 /evidence=not_experimental
 /product="Ylr454wp"
 /protein_id="AAB64718.1"
 /db_xref="GI:717075"
 /translation="MMFPIINVLLYKMLFVAVPLMSCKTLRLGINTWLNFLKE
 IGLSLSDGTVLKSVFAPVPERKLFKGLRIDSCKSSNDLAKELPREESTFETEP
 EDNGGFSIKLISLSQYWLNGVTIILEDQVANNDDTIKRFEPFLSINDSKITLRF
 DSFLRLKLNNGOTTIADALFIVNTNLIGBINPLKDGQVGLDKLGLNLPNNILN
 LFLKLNVDLMSNEKRLQRLADTKNEBKEDLAKMDLVYAMEKEDVRIKPLKE
 MNTVTVKIQIDPLTNHBLGMMKYISYVLSVINENFRFRMEKGYTLIFEEER
 DSPEFSIIMARENIYLNLRKSHAKOLKTIIEIPVSI FGETNLFSQKPLSNHLK
 AKLENAITINIGNISLTITDMDPVNISFKELSNIKYFTSSCPNKTKLKNSHYK
 LTRRVLDFPKFCFLINMKTLDLPKVIINDKDLIIGKSVFISHSKRYTLGN
 NMBEKEETQHLFYESHNVVELDMQLQHLIKQKQKHTLKVDSIAIEKVQLPDI
 LGSANADITLMDLSELPTMVLSELVNLSQLANVENEFKEFEYFAANLQMK
 ACSNNAKCLQKQKEILPSDFMFOQLPDFDIYIKINRISITGLASVMPDPVSSV
 DSOSSKDLIDGKLRKYNCTVEKQLALFDGKQMNKIGSSNATWRSGLNFSKDN
 KQNPNNKSSIALDDISTSDATEVNLNMINLVNDITTSITGETPEVSEBELSTKTVS

gene
 CDS
 11123..11737
 /gene="YLR456W"
 11123..11737
 /gene="YLR456W"
 /codon_start=1
 /evidence=not_experimental
 /product="Ylr456wp"
 /protein_id="AAB64720.1"
 /db_xref="GI:717073"
 /translation="MKNEQIQPKDLRLIKSSKYVAVATCSSNIPSVLAHYFVS
 AETPKHSEYIEIDONDIITFVEKSTFPPNNAVSNPVALLPDMWTKALTKRKS
 VHSKDDSEFVSESEKFPNNPRLDNGSLNAYSATINCIADIVAPNSESTYRRLL
 TVNPDAIDFIIIGEDTAIKVNIQIKVSDMENNSTIGQVQV"
 complement(11865..12824)
 /gene="NBPI"
 /db_xref="S288C:10001234"
 complement(11865..12824)
 /gene="NBPI"

repeat_unit
 gene
 CDS
 9682..10596
 /gene="YLR455W"
 9682..10596
 /gene="YLR455W"
 /codon_start=1
 /evidence=not_experimental
 /product="Ylr455wp"
 /protein_id="AAB64719.1"
 /db_xref="GI:717074"
 /translation="WTKDIRTGDVLVCKVGSFPMPVAVFPORLLRNDVYRRKSNVCV
 AVCFNDPTTYWEQPSRLKELDQDSIHNFILSHSKNANQRELVNAYKEAKPPDDPNV
 LOEKFEENRSLDLKAFKESGSKIYVAEDFVGATKYVNRKKSISIKEDPENQK
 SNEESKPNIKPSKKKRPYANSCKSKSNKKYALDYSRVEISQLPFRRIORLQI
 RPTPTHEIKETHEILLNRIYENSDPDKPFDKALRSKHKLLKALVANDPDLGEFH
 PLCKRILSNADLITELKKEKQLPPT"

tRNA
 gene
 complement(8124..8197)
 /gene="tI(AAU)12"
 complement(8124..8197)
 /gene="tI(AAU)12"
 /product="tRNA-Ile"
 /note="codon recognized: AUU"
 complement(8357..8685)
 /note="delta element YLRdelta27"
 9682..10596
 /gene="YLR455W"
 9682..10596
 /gene="YLR455W"
 /codon_start=1
 /evidence=not_experimental
 /product="Ylr455wp"
 /protein_id="AAB64719.1"
 /db_xref="GI:717074"
 /translation="WTKDIRTGDVLVCKVGSFPMPVAVFPORLLRNDVYRRKSNVCV
 AVCFNDPTTYWEQPSRLKELDQDSIHNFILSHSKNANQRELVNAYKEAKPPDDPNV
 LOEKFEENRSLDLKAFKESGSKIYVAEDFVGATKYVNRKKSISIKEDPENQK
 SNEESKPNIKPSKKKRPYANSCKSKSNKKYALDYSRVEISQLPFRRIORLQI
 RPTPTHEIKETHEILLNRIYENSDPDKPFDKALRSKHKLLKALVANDPDLGEFH
 PLCKRILSNADLITELKKEKQLPPT"

/note="YLR457C; Weak similarity in central portion of the protein to tropomyosin alpha chain (Swiss Prot. accession number P18442) and kinesin heavy chain (Swiss Prot. accession number P21613)"

/codon_start=1
/evidence=not_experimental
/product="Nbp1d"
/protein_id="AAB64721.1"
/db_xref="GI:717076"
/db_xref="SGD:L0001234"
/translation="MLKSVQGLMKDFGIRDDRRREYSGLSDEVKRSALSRKQMR
PGKSVLRKPRVTRKTEKIRTRKTRKRLKIRQIRDIVSNENNSKQNY
CGDMTRILKRSQGRPSYMDTAKSRILRSDFKXISELYNFORISELSGSSDG
SSGDRNOSLYDRILOROKKRDKEKALSKESLOEALYNSNEKRYLIEDLID
SNHPSYTKSRRTSMNLARENDEKPKLIDSPSRIRTSLSFTSSPMKYNRDGI
PEMOPLOENISPACTPPRSHREKEDETLSPISVDSSYLS"

gene
13387.14571
/gene="CDC91"
/db_xref="SGD:L0000290"
13387.14571
/gene="YLR459W"
/note="YLR459W"
/codon_start=1
/evidence=not_experimental
/product="Cdc31p"
/protein_id="AAB64722.1"
/db_xref="GI:717072"
/db_xref="SGD:L0000290"
/translation="MDSTALKVALACIARLAVNSLFPSLQOOLDSVEFSTPVSFR
SLOEGIYLRNNIOVYNGVHNPILIFELFNSDRSLISLYALIGLAVQULEV
TAFPNLKLKWLPGILVAVNPILSLFTNPALISSYCLIAEENVLSS
VMSISGVSVYRILLILPLGLMKSMQRILSAVISLILLLFVSILSGSSMS
ELTOYVSGIITPEKYPNIGLWVYPRIPERPAVNIPIAVITPPTLPHYK
QPIVAFILICGIVITKPYPSIGDAPFFSFLPFTPLRGHURIPYISALFLHAILV
APIFYHLWLVLSGNSNPFYALSILVALAISILVLDLWMLRIEYDNGIPIVFKLTV
QI"

gene
complement(15810..16940)
/gene="YLR460C"
complement(15810..16940)
/gene="YLR460C"
/note="Similar to hypothetical 40.1 kD protein in HMR 3'

CDS
13387.14571
/gene="CDC91"
/db_xref="SGD:L0000290"
13387.14571
/gene="YLR459W"
/note="YLR459W"
/codon_start=1
/evidence=not_experimental
/product="Cdc31p"
/protein_id="AAB64722.1"
/db_xref="GI:717072"
/db_xref="SGD:L0000290"
/translation="MDSTALKVALACIARLAVNSLFPSLQOOLDSVEFSTPVSFR
SLOEGIYLRNNIOVYNGVHNPILIFELFNSDRSLISLYALIGLAVQULEV
TAFPNLKLKWLPGILVAVNPILSLFTNPALISSYCLIAEENVLSS
VMSISGVSVYRILLILPLGLMKSMQRILSAVISLILLLFVSILSGSSMS
ELTOYVSGIITPEKYPNIGLWVYPRIPERPAVNIPIAVITPPTLPHYK
QPIVAFILICGIVITKPYPSIGDAPFFSFLPFTPLRGHURIPYISALFLHAILV
APIFYHLWLVLSGNSNPFYALSILVALAISILVLDLWMLRIEYDNGIPIVFKLTV
QI"

Query Match 32.3%; Score 359; DB 8; Length 29634;
Best Local Similarity 58.9%; Pred. No. 1.6e-69;
Matches 658; Conservative 0; Mismatches 450; Indels 9; Gaps 2;

5 CAGTTCCACCACTCAAAAAGCCGTCATCATTTGAAGTGACAAAGCTGTGTTAAACAG 64
16930 CAATTCAGAAACCAATGAAGGCTGTCATTTGAAGACGTAAGCGTTGTTAAAGAG 16871

65 ATGCTCAGTTCCAGAAATTAAGAGAGGTAACGCTTGTAAGGTGAGGCTGTGCTG 124
16870 GCATTCACCAATTCCTGAATTAAGAGAGAGTTCGATTAAGTAAACACTGCTGTGCTG 16811

125 GTAACCACTGATTTGAAGCATATGCTTAATAGTTGTCGAGAAAGTCAATTCAG 184
16810 GTAAACCCCACTGATTTGAGGACACATTTGACTAAGATCGGCTCAAGGATCTATTCG 16751

185 GATGTACATTCGCTGTACAGTTGTCAACTTGACCAAAATCTAGTACT--GACTTGA 241
16750 GATGTATGCTGCTGGCCAAATTTGTCAAATTTGGGCCCACTGTCATCTTAAGACTTTT 16691

242 AGGTGGAATACCGGTTTCGTTTGTTCACAGGTCCTTCCAAACAGATCTTAAAAAG 301
16690 CTATCGGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 16631

302 GTGCACTTCTGAATATGACAGGTTTATCCACTTGTGTTTAAAC-----GAGTAAT 355
16630 GTGCTTTTCTGAATATTTCTGATTTTCACTGTGTGTTGCTTCAAAATCAACCAATGAG 16571

356 TAACTCACTCAACTGCTGATGAATTTTGAAGGCCCTGTGAAGAACTTGAATCTGCTG 415
16570 TCAATTTTGGGTGAGATGTTCTACTGCGCGCCCTGTCAAGTCTTGAAGGTTAG 16511

416 CATCATTCGCAAGTTTCGTTGACCAACTGCTGTTAGTTGTGTCATCATCTGGGCTCAA 475
16510 CCATCATTCGCAAGTTTCGTTGACCAACTGCTGTTAGTTGTGTCATCATCTGGGCTCAA 16451

476 AATGGAATGACCACTCAATCAACCCCGCAACATCACTCATATTAATTTGGGCTG 535
16450 AACTGGAATGACCACTCAATCAACCCCGCAACATCACTCATATTAATTTGGGCTG 16391

536 CTACAGCAGTGGGTCACAACTTAATCAAGTTGCCAAACATATCAATCTTATATAAGA 595
16390 CAATGCAAGTGGGTCACAACTTAATCAAGTTGCCAAACATATCAATCTTATATAAGA 16331

596 TTGTAAGCTGTTCTCTTAAAGATGAAAGCTTTTAAAGCTTAATGCTGATGATGATG 655
16330 TCATTTGTTGCTTCTCTTAAAGATGAAAGCTTTTAAAGCTTAATGCTGATGATGATG 16271

656 TCTTGAATCATGATGACAGCGCTTATTAAGCAGATCAAAATCGAATATCAAACTG 715
16270 TATTTGATTAATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 16211

716 AACATGTTATGACGCTGTGGAAAGCAAGATGATATCCGAGGCTTATTAAGTACAG 775
16210 CGTATTTAGTGAAGTGTGCGAATCAAGATGATGATGATGATGATGATGATGATGATG 16151

776 CAGATAGTCTACCTGCACTTATTAAGAGGTTCCAAATGACCAATGAAAGCAATTCCTG 835
16150 CGATTAAGCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 16091

836 AAGAAATCAGAAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 895
16090 AAGAAATCAGAAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 16031

896 AAGAAATCAGAAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 955
16030 ATGAAGTCAATTTGAAAGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 15971

956 TTAATTCGTTAAGTTTAAATCAACCTTAACCAAGGATATCAATCAATTAATTAATTA 1015
15970 TTAATTCGTTAAGTTTAAATCAACCTTAACCAAGGATATCAATCAATTAATTAATTA 15911

1016 TTAAGTTTTCAGCAACCGCTTAAGTATGATGATGATGATGATGATGATGATGATGAT 1075
15910 TTAAGTTTTCAGCAACCGCTTAAGTATGATGATGATGATGATGATGATGATGATGATG 15851

1076 GTAAACCAAAATGTTAATGATGATGATGATGATGATGATGATGATGATGATGATG 1112
15850 GTAAACCAAAATGTTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 15814

RESULT 2
SCYNL34C/c 1944 bp DNA linear PLN 11-AUG-1997
LOCUS SCYNL34C
DEFINITION S.cerevisiae chromosome XIV reading frame ORF YNL134C.
ACCESSION Z71410.1
VERSION Z71410.1 GI:1302075
KEYWORDS
SOURCE
ORGANISM
Saccharomyces cerevisiae.
Saccharomyces cerevisiae.
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE
1 (bases 1 to 1944)
Mallet, L., Busseret, F. and Jaquet, M.
JOURNAL
Unpublished
AUTHORS
2 (bases 1 to 1944)
MIPS.
TITLE
Direct Submission
JOURNAL
Submitted (29-APR-1996) Data collected by MIPS on behalf of the
European yeast chromosome XIV sequencing project. MIPS at the
Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a D-82152
Martinsried, FRG; E-mail: Mewes@mips.emblnet.org
FEATURES
source
1..1944

REFERENCE
AUTHORS 5 (bases 17182 to 18759)
Field,J., Vojtek,A., Ballaster,R., Bolger,G., Colicelli,J., Ferguson,K., Gerst,J., Katoaka,T., Michaeli,T., Powers,S., Riggs,M., Rodgers,L., Wieland,I., Wieland,B. and Wiegler,M.
Cloning and characterization of CAP, the *S. cerevisiae* gene encoding the 70 kd adenyl cyclase-associated protein
Cell 61 (2), 319-327 (1990)

JOURNAL
MEDLINE 90235281
PUBMED 2184942
REFERENCE 6 (bases 22326 to 22667)
Wiederrecht,G., Brizuela,L., Elliston,K., Sigal,N.H. and Sietekka,J.J.
FKB1 encodes a nonessential FK 506-binding protein in *Saccharomyces cerevisiae* and contains regions suggesting homology to the cyclophilins
Proc. Natl. Acad. Sci. U.S.A. 88 (3), 1029-1033 (1991)

JOURNAL
MEDLINE 91126049
PUBMED 1704127
REFERENCE 7 (bases 22326 to 22667)
Heitman,J., Movva,N.R., Hiestand,P.C. and Hall,M.N.
FK 506-binding protein proline rotamase is a target for the immunosuppressive agent FK 506 in *Saccharomyces cerevisiae*
Proc. Natl. Acad. Sci. U.S.A. 88 (5), 1948-1952 (1991)

JOURNAL
MEDLINE 91156723
PUBMED 1705713
REFERENCE 8 (bases 22326 to 22667)
Koltin,Y., Faucecette,L., Bergsma,D.J., Levy,M.A., Cafferty,R., Koser,P.L., Johnson,R.K. and Livi,G.P.
Rapamycin sensitivity in *Saccharomyces cerevisiae* is mediated by a papydylin-prolyl cis-trans isomerase related to human FK506-binding protein
Mol. Cell. Biol. 11 (3), 1718-1723 (1991)

JOURNAL
MEDLINE 91115524
PUBMED 1596117
REFERENCE 9 (bases 19039 to 20496)
Boguta,M., Dmochowska,A., Boruck,P., Wrobel,K., Gargouri,A., Lazowska,J., Slonimski,P.P., Szczesnaki,B. and Kruszezka,A.
NAMS nuclear suppressor of mitochondrial ochre mutations in *Saccharomyces cerevisiae* codes for a protein homologous to S4 ribosomal proteins from chloroplasts, bacteria, and eucaryotes
Mol. Cell. Biol. 12 (1), 402-412 (1992)

JOURNAL
MEDLINE 92107194
PUBMED 1729612
REFERENCE 10 (bases 1 to 43481)
Mallet,L., Bussereau,F. and Jacquuet,M.
A 43.5 kb segment of yeast chromosome XIV, which contains MPA2, MEP2, CAP/SRV2, NAMY, FKB1/FPPI, MOW2 and CPPI, predicts an adenosine deaminase gene and 14 new open reading frames
Yeast 11 (12), 1195-1209 (1995)

JOURNAL
MEDLINE 96109932
PUBMED 8619318
REFERENCE 11 (bases 29208 to 29663)
Honlinger,A., Kuebrich,M., Moczko,M., Gaertner,F., Mallet,L., Bussereau,F., Eckerskorn,C., Lottspeich,F., Dietmeier,K., Jacquet,M. and Pfanner,N.
The mitochondrial receptor complex: Mom2 is essential for cell viability and directly interacts with preproteins
Mol. Cell. Biol. 15 (1995) In press

JOURNAL
REFERENCE 12 (bases 7894 to 9390)
AUTHORS Martini,A. and Andre,B.
MEP2
Unpublished
13 (bases 1 to 43481)
Jacquet,M.
Direct Submission
Submitted (29-NOV-1994) Michel Jacquet, Institut de Genetique et Microbiologie, Universite, Paris-Bud, Laboratoire Information Genetique et developpement, Orsay, 91405, FRANCE
On Jun 14, 1995 this sequence version replaced gi:854489.
Location/Qualifiers
1..43481
/organism="Saccharomyces cerevisiae"
/strain="alpha S288C"

COMMENT
FEATURES
source

CDS
/db_xref="taxon:4932"
/chromosome="XIV (43.5kb fragment)"
/clone="alpha 14-12"
/clone_lib="from Peter Philippsen"
complement(348..1112)
/citation=[10]
/codon_start=1
/label=JTD254
/product="cytoskeleton-associated protein (putative)"
/protein_id="CAA6878.1"
/db_xref="GI:854490"
/db_xref="SWISS-PROT:P53904"
/translation="MVRVIESELVREKELPNSLKURQFKRLYHTGVPEDEMEIVKRYDNKEIYSTKGGAGNEDENAFLEGEELIVVTDNKSQISNOLATQAEI PSMVIESEEDYTRDQSVLRKRAHGVRNAQSORALAKQDEAVARFOLTRAIG RHCRTVDGSHPREALIRTVGPLPLDMGTGCGVFPFAAGKRDGRINGTTLFGPVAP GHGSEVRPAVEILSKDESAEVEDVDHSDDEI"
join(1381..1398,1519..1824)
/gene="orf2"
join(1381..1398,1519..1824)
/gene="orf2"
/note="putative member of the Sm protein family"
/citation=[10]
/codon_start=1
/label=JTA107
/protein_id="CAA6879.1"
/db_xref="GI:854491"
/db_xref="SWISS-PROT:P53905"
/translation="MHQHQSKKFEKPEKBAIDLAKYDQSKIRVLMGGKLVIGVLKGYQLNMLVDDTVVEYMSNPDDNNTELISKNRKLGTVIRGTILVLSABGSDV LYMOK"
1398..1518
/gene="orf2"
2156..2458
/gene="orf3"
2156..2458
/gene="orf3"
/citation=[10]
/codon_start=1
/label=JTB100
/protein_id="CAA6880.1"
/db_xref="GI:854492"
/db_xref="SWISS-PROT:P53906"
/translation="MSNTKHTSHHMLRIILRLIFILMIFRNSVSRKMFQEL LPREYKQNSNSVSNRNPSSIFSNLVDPDVNVMDTRLPFLFPFIITIPMV"
2855..2971
/gene="wfa2"
2855..2971
/gene="wfa2"
/function="induces mating processes on alpha responsive cells"
/citation=[2]
/citation=[10]
/codon_start=1
/label=JTB38
/product="mating pheromone a-factor 2"
/protein_id="CAA6881.1"
/db_xref="GI:854493"
/db_xref="SWISS-PROT:P34166"
/translation="MQPIITASTQATQKDSSEKQNYIIGLFWDBACVIA"
complement(3261..5483)
/gene="orf5"
complement(3261..5483)
/gene="orf5"
complement(3261..5483)
/gene="orf5"
/citation=[10]
/codon_start=1
/label=JTD740
/protein_id="CAA6882.1"
/db_xref="GI:854494"
/db_xref="SWISS-PROT:P53907"
/translation="MSSSIFEMTIDHGMSTGKLQSTASSISMGDFELCSSTTNS ILDSPLPVTENHDSITDINTINMEIIVEQSGVDVADKNVLVYCIDPYVPVPCY DEANPSKIRIRYIEHCRPCLTSVSPSYTSPVEHYTVVSKMEKLSPEFASSRLMN

NPLQNSTOINFIYSDIDSLTRHIAKVRGDMVDHSHSKTASDRHHSRLINATTT
 KSTYQDKYDKERI CCEIADDEHFLSDELSYSLSQCKVLPIDYSRDLVLMR
 CEGQOPLVOPSHVDELIYAMYLNMGISLSLDELEMPYRSEVPRRPRARFPOH
 NKNNKNSKRONSDRNDRSRSHSLLRSHSISVYTRKATNSRSPKSSRSLSLIP
 SVSDASHSDSPNDGSLAGSOTDFCGLFSTKLNPFPTDSSSKNSRSMNDGOKR
 SKELNVOERIDNDGSTTNTNTSVLSSTPSHVSILTQOTSHEHPRSRNSNPIDP
 LRCDRSVLKINNEPEFYEGTGRSTISNSAERTREKLNKNGDNVDNDEDDDEDA
 NYEDEDDEVEDDINRLVYLDERSRESRYENADVEYGTIFSTYSFVTCGLNPFK
 LSLKCNKQSSNDSSKWPATOLVSRKRYKOSLRICIKPLTEDHPWVGKIKPRPALP
 AFETNNPPIRVYSGEDSTDLHVKNHYLKRYIVGSCFLTGSKLHFHFDRTDLNLT

Query Match 32.1%; Score 357.8; DB 8; Length 43481;

Best Local Similarity 58.8%; Pred. No. 2.9e-69;

Matches 658; Conservative 0; Mismatches 452; Indels 9; Gaps 2;

4 TCAGTTCCAAACCACTCAAAAAGCCGTCATCATTTGAAGTGAACAAGCTGTTGTTAAACA 63
 24013 TCGATTCCAGAAACCATGAAAGCCGTTGTCATTTGAATGCGAAGCTGTAGTCAACAG 23954
 64 GATGTCCTCACTTCAGAAATTAAGAGAGGAGTACAGCTTGTGAAGGTTGAGCTGTGCT 123
 23953 GACATTCCTCAATTCCTGTAATTAAGAAAGATTTGTTCTAATTAAGCTGTGCGCTTGCC 23894
 124 GGAACCCCACTGATGGAAGCATTTGCTTAAAGTTGTCGAGAAGGTTCAATTCCTA 183
 23893 GGTAAACCTTACCGATTTGGAACATATTGATTTCAAGATTGCTCTCAAGGTGCTCTTA 23834
 184 GATGTCACATTCCTGTGTACAGTTGTCAAACTTGACCAATGCTAGTACTG---ACTTG 240
 23833 GGTGTCGATGACACCGCCGCAAAATCTGAAGTTGGGCCCAATGTGATGCTGACGCTTT 23774
 241 AAGGTTGAGATACCGGTTTGGTTTGTGACGGTGTCTCCAAACAGATCTGTAATAAT 300
 23773 GCCATTGGTATATACATTATTTGAGGTTATTCACGGTGTCTCAAGTGTGCTCCCTCAAC 23714
 301 GGGCATTTGCTGAATATGCGAGGGTTATCCACTTGTGTTTACA-----GAGTAAC 354
 23713 GGTGCTTGTCTAGTACTTGTCCATTTCAATCCGACATGCTGTTAAACGACCCAGAGAG 23654
 355 TTAACCTACTCACTGCTGTGTAATTTCTGAAGCCCTGTGAAGCACTTGAATCTGCT 414
 23653 TTTAGATTGTGGGTAAGACAGCTACGAGAAAGCCCGTAAATCTTTAGAAAGGGGCA 23594
 415 GCATCATTTGCGAGTTTGTGTAACAATGCTGTGTGTTAGTTGTGTCATCACTTGGGCTCA 474
 23593 GATTCCTCCAGTCTCATTTAGCACGCTGTGATGATCTTACACATAGTTTGGGCTTG 23534
 475 AAAATGGAATGGACCATCTACCCCGCAACATACATCATTTATTTGGGGTGT 534
 23533 GACATGACATGAAAGCCCTCCAAAGCGCAAGAGATTAACCCATTTATTTGGGGTGT 23474
 535 GCTACAGCAGTGGGTCAACAACATACTCAAGTTGCCAACAATATCAATGCTTACTAAG 594
 23473 GCGACAGCTGTGGCCAGATGCTTATTCATTTGGCAAAAAAATCAACGTTTCAAGAG 23414
 595 ATTGTAACTGTGCTTCTTAAAGCATGAAGCTTTTAAAGCTTTATGTGTGCTGATGAT 654
 23413 ATCATGCTGCTTCTTCTGTAACAAGAAAAATGTTGTAAGAGTACGCTGACATGAGA 23354
 655 GTCCTTGACTATGATGAGGAGGTTATGAGCAGATCAATGGAATGCAAAACCTG 714
 23353 CTTTGTGACTACACAGATGCTGAGTTATGAAACAGTAAAGAAAGATCAACACATTT 23294
 715 CAACATGTTATTAAGCCTGTGGAGAGAGATGATATCCCGAGGCTTATTAAGTACA 774
 23293 CTTTACTGTGTGAGCTGTGCTCCACACAGAAACATATTCACAGGTGTCAATATGTC 23234
 775 GCAATATGCTTACCTGCGACATTTAGAAAGTGTTCATGACCATTTGAAGCATTTCT 834
 23233 GCTGAATGACTTAACGCTACGCTGTTCATTTAACCCTTTAAACGAAAAAGATATCAAG 23174
 835 GAAGAATTCAGAAAGTAAATGTAATTAATGATTAATCTTGTGTATGCTGATCTGCT 894

Db 23173 GAGGAAGACAGAGGCAAAACGTCAGTATTAAGAACCTTCTATATTGATAGAGCT 23114
 QY 895 CAGAAATTCATTTGGGTGCAACAAGATTTCTGCTAGTCAGAAATATCATGAAGCCACA 954
 Db 23113 AACGAGCTCCATTTGGCAGCGTTTACTTGTCCAGAGACCTGAAATVACAGAGAGCCGCC 23054
 QY 955 GTTAAATTTGTTAAGTTTAAATCAACACCTTAAACAGGAGATATCATATATGAT 1014
 Db 23053 ATAAAAATTTATTAAGTTATCAATCAAAAATCAATGATGGAATCCACACATCCCA 22994
 QY 1015 ATTAAGTTTTCAGACAGCGCTTATGATGATGCCAGCTCTCAGTGAAGTATTAAGA 1074
 Db 22993 GTGAAGTTTACAGAACCGGTTAGATGATATCCACAGTTACTTGATATTTAAGCAC 22934
 QY 1075 GGTAAACCAAAATGTTAAGTATGTCGACAGTTATTA 1113
 Db 22933 GGGAGAAATTCGGCGAAAGTTGGTGGCGTCTGAAA 22895
 RESULT 4
 SCCHRII/c
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 SCCHRII
 S.cerevisiae chromosome III complete DNA sequence.
 X59720.2 S43845 S49180 S58084 S93798
 X59720.2 GI:14588895
 chromosome.
 baker's yeast.
 Saccharomyces cerevisiae
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetidae; Saccharomycetaceae; Saccharomyces.
 1 (bases 1 to 316613)
 Rad,M.R., Lutkenltichen,K., Xu,G., Kleinhaus,U. and Hollenberg,C.P.
 The complete sequence of a 11,953 bp fragment from Clg on
 chromosome III encompasses four new open reading frames
 yeast 7 (5), 533-538 (1991)
 91377317
 1897318
 2 (bases 1 to 316613)
 Bileau,N., Fremaux,C., Hebrard,S., Menara,A., Aigle,M. and
 Crouzet,M.
 The complete sequence of a 10.8kb fragment to the right of the
 chromosome III centromere of Saccharomyces cerevisiae
 yeast 8 (1), 61-70 (1992)
 92254505
 1580102
 3 (bases 1 to 316613)
 Wilson,C., Bergantino,E., Lanfranchi,G., Valle,G., Carignani,G. and
 Frontali,L.
 A putative serine/threonine protein kinase gene on chromosome III
 of Saccharomyces cerevisiae
 yeast 8 (1), 71-77 (1992)
 92254506
 1580103
 4 (bases 1 to 316613)
 Bentl,P., Chanet,R., Fabre,F., Faye,G., Fukuhara,H. and Sor,F.
 Sequence of the sup61-RAD18 region on chromosome III of
 Saccharomyces cerevisiae
 yeast 8 (2), 147-153 (1992)
 92221691
 1561837
 5 (bases 1 to 316613)
 Bolle,P.A., Gillinguet,V., Berben,G., Dumont,J. and Hilger,F.
 The complete sequence of K3B, a 7.9 kb fragment between PGK1 and
 CRY1 on chromosome III, reveals the presence of seven open reading
 frames
 yeast 8 (3), 205-213 (1992)
 92245758
 1574926
 6 (bases 1 to 316613)
 Sor,F., Charet,G., Fabre,F., Faye,G. and Fukuhara,H.
 Sequence of the HMR region on chromosome III of Saccharomyces
 cerevisiae
 yeast 8 (3), 215-222 (1992)
 JOURNAL

JOURNAL MEDLINE 92245759
PUBMED 1574927
REFERENCE 7 (bases 1 to 316613)
AUTHORS Oliver,S.G., van der Aart,Q.J., Agostoni-Carbone,M.L., Aigle,M.,
Alberghini,L., Alexandraki,D., Antonic,G., Anwar,R., Ballesta,D.P.,
Bent,P. et al.
The complete DNA sequence of Yeast chromosome III
Nature 357 (6373), 38-46 (1992)

JOURNAL MEDLINE 92244356
PUBMED 1574125
REFERENCE 8 (bases 1 to 316613)
AUTHORS Skala,J., Purnelle,B. and Goffeau,A.
The complete sequence of a 10.8 kb segment distal of SUP2 on the
right arm of chromosome III from Saccharomyces cerevisiae reveals
seven open reading frames including the RVS161, ADP1 and PGK genes
Yeast 8 (5), 409-417 (1992)

JOURNAL MEDLINE 92327849
PUBMED 1626432
REFERENCE 9 (bases 1 to 316613)
AUTHORS Wilson,C., Grisanti,P. and Frontali,L.
The complete sequence of a 6146 bp fragment of Saccharomyces
cerevisiae chromosome III contains two new open reading frames
Yeast 8 (7), 569-575 (1992)

JOURNAL MEDLINE 92397594
PUBMED 1523889
REFERENCE 10 (bases 1 to 316613)
AUTHORS Scherens,B., Messenguy,F., Gigot,D. and Dubois,E.
The complete sequence of a 9,543 bp segment on the left arm of
chromosome III reveals five open reading frames including
glucokinase and the protein disulfide isomerase
Yeast 8 (7), 577-585 (1992)

JOURNAL MEDLINE 92397595
PUBMED 1523890
REFERENCE 11 (bases 26740 to 32076)
AUTHORS DeGoor,E., Debrabandere,R., Keyers,B., Voet,M. and Volckaert,G.
Nucleotide sequence of D108, a BamHI fragment on the small-ring
chromosome III of Saccharomyces cerevisiae
Yeast 8 (8), 681-687 (1992)

JOURNAL MEDLINE 93070606
PUBMED 1441748
REFERENCE 12 (bases 169581 to 171116; 171683 to 172169)
AUTHORS Agostoni Carbone,M.L., Panzeri,L., Muzi Falconi,M., Carcano,C.,
Plevani,P. and Lucchini,G.
Nucleotide sequence of 9.2 kb left of CRY1 on yeast chromosome III
from strain AB972: evidence for a Ty insertion and functional
analysis of open reading frame YCR28
Yeast 8 (9), 805-812 (1992)

JOURNAL MEDLINE 93070619
PUBMED 1332309
REFERENCE 13 (bases 1 to 315338)
AUTHORS Valle,G.
TA-repeat microsatellites are closely associated with ARS consensus
sequences in Yeast chromosome III
Yeast 9 (7), 753-759 (1993)

JOURNAL MEDLINE 93377412
PUBMED 8368009
REFERENCE 14 (bases 1 to 315338)
AUTHORS Slonimski,P.P. and Brouillet,S.
A data-base of chromosome III of Saccharomyces cerevisiae
Yeast 9 (9), 941-1029 (1993)

JOURNAL MEDLINE 94091061
PUBMED 8266725
REFERENCE 15 (bases 1 to 315339)
AUTHORS Rodriguez-Cousino,N., Lili,R., Neupert,W. and Court,D.A.
Identification and initial characterization of the cytosolic
protein Ycr77p
Yeast 11 (6), 581-585 (1995)

JOURNAL MEDLINE 95373282
PUBMED 7645349
REFERENCE 16 (bases 1 to 316613)
AUTHORS MIPS.
Direct Submission
Submitted (16-MAR-1992) MIPS, D-8033 Martinsried, FRG. Data

REMARK	collected by MIPS on behalf of the European Yeast Chromosome III Sequencing project
REFERENCE	replaced by [13]
AUTHORS	17' (bases 1 to 314957)
JOURNAL	Jimenez, A.
REMARK	Direct Submission
REFERENCE	Submitted (28-DEC-1992) MIPS, D-8033 Martinsried, FRG. Data collected by MIPS on behalf of the European Yeast Chromosome III Sequencing project. Update originating from A. Jimenez
AUTHORS	replaced by [14]
TITLE	18 (bases 1 to 316613)
JOURNAL	Louis, E.J.
REMARK	Direct Submission
REFERENCE	Submitted (25-JUN-1993) MIPS, D-8033 Martinsried, FRG. Data collected by MIPS on behalf of the European Yeast Chromosome III Sequencing project. Update originating from E.J. Louis
AUTHORS	revised by [18]
TITLE	19 (bases 1 to 316613)
JOURNAL	Louis, E.J.
REMARK	Direct Submission
REFERENCE	Submitted (14-FEB-1995) MIPS, D-8033 Martinsried, FRG. Data collected by MIPS on behalf of the European Yeast Chromosome III Sequencing project. Update originating from E.J. Louis
AUTHORS	Revised by [20]
TITLE	20 (bases 1 to 316613)
JOURNAL	Gromadka, R.
REMARK	Direct Submission
REFERENCE	Submitted (29-JAN-1996) R. Gromadka, Protein Biosynthesis, Institute of Biochemistry and Biophysics, Pawinskiego 5A, Warsaw, Poland, Electronic Mail Address: robert@psd.tlb.waw.pl
AUTHORS	revised by [21]
TITLE	21 (bases 1 to 316613)
JOURNAL	MIPS.
REMARK	Direct Submission
REFERENCE	Submitted (17-JUN-2001) MIPS Yeast Genome Database, GSF - Ingolstaedter Landstrasse 1 D-85764 Neuherberg, Germany,
AUTHORS	mips-yeast-adm@gsf.de
TITLE	Resequencing project, achieved by the joint effort of G. Valles and G. Voickaerts laboratories. Munich information center for protein sequences.
JOURNAL	
REMARK	
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
COMMENT	On Jul 3, 2001 this sequence version replaced g1:1907116. The contig sequence which has been released to the file server in 1992 has been subject to a resequencing project, achieved by the joint effort of G. Valles and G. Voickaerts laboratories. The following lines present a summary of the altered entities. The resequenced chrl1i contig is 316613 bp with GCG-check: 429 GENETIC ENTITIES valid no longer: YCL012w (YCL014w + YCL012w -> YCL014w / C_B1367 + C_C231->C_A1636), PEP Y, ORF in identical form now C-terminal part of YCL014w. (A frame shift in the nt sequence of YCL014w leads to an elongated polypeptide.) This elongated polypeptide encoded by the current YCL014w resembles the Bbdp polypeptide from the EMBL-entry (Accession:SC175801;ID:U17580, total length 1636 aa) YCL006c, PEP Y, ORF obsolete, is not an ORF in current Chromosome III contig. YCR062w (YCR061w + YCR062w -> YCR061w / C_A583 + C_B120 -> C_B631), PEP Y, ORF in identical form now C-terminal part of YCR061w. YCR066w-a (YCR068w + YCR068w-a -> YCR068w / C_C429 + C_A145 -> C_A507), PEP Y, ORF in identical form now C-terminal part of YCR068w. YCL103c (C_F111, former ORF in region corresponding to coord. 307255-307581(C)) is obsolete, open reading frame is destroyed by single nt exchange resulting in a stop codon after 45 aa. NEW GENETIC ENTITIES: YLC26c-b (C_E193(YCL027c-a)), PEP Y, new ORF YCL021w-a (C_C125), PEP Y, new ORF ALTERED GENETIC ENTITIES: YCL076w, PEP Y, peptide elongated N-terminally by 51 aa YCL074w, PEP Y, 3 aa exchanged YCL073c, PEP Y, 4 aa exchanged YCL066c (C_F190->C_E260), PEP Y, peptide elongated N-terminally by 70 aa YCL065w, PEP Y, 1 aa exchanged


```

misc_feature <34..>1008
              /note="similar to Saccharomyces cerevisiae ORF YCR102c [
              /evidence=not experimental
BASE COUNT   330 a   206 c   239 g   306 t   2 others
ORIGIN
Query Match      22.7% Score 252.4; DB 11; Length 1083;
Best Local Similarity 55.7%; Pred. No. 1.1e-45;
Matches 596; Conservative 2; Mismatches 443; Indels 29; Gaps 5;

QY 1 ATGTCAGTTCCAAACCACTGAAAGCCGTCATCTTGAAGAGTGAACAGCTGTGTTAA 60
   |||||
DB 16 ATGTCAGTTCCAAACCACTGAAAGCCGTCATCTTGAAGAGTGAACAGCTGTGTTAA 75
   |||||
QY 61 ACAGATGTCAGTTCCAAATTAAGAGGCTGACGCTTGTGAGGTTAGAGCTGTT 120
   |||||
DB 76 GAAAACGTTCTTACCACTCTCAAGAGGCTCAAGTGTAGTCAAGCGCTTGACGCT 135
   |||||
DB 121 GCTGTAACCAACCTGATGAAAGCATATGCTTATAGATGATGTCAGAGAGTTCAATT 180
   |||||
DB 136 GCTGTTATCCACGAGATGGAAGCATGATGATTAACCTTGTCACAAAGCTCAATT 195
   |||||
QY 181 CTAGAGATGACATGCTGCTGACAGTTGCAAACTTGAACCAATGCTAGTGAAGT 240
   |||||
DB 196 TCAGAGATGACGCTGCGGACAGATGTAAGTTAGGGGCTAACGTTAAATGTCGCT 255
   |||||
QY 241 AAGGTTGAGATACCGGTTCCGTTTGTGTCACGCTGCTCCCAACAGATCTCTAAAT 300
   |||||
DB 256 GAG---GGTATGTTGTGCGGATTCGACATGCGATCTGTCMAACATCCGAAAT 312
   |||||
QY 301 GGTGATGTTGCTGAATATGTCAGAGGTTATTCACCTTGTGTTTACAGAGTAACT 360
   |||||
DB 313 GGTGCTTTGCGAATATGTTGCAATGATTTGAGTTGCTTGTGAAATTTGCAATCGAGT 372
   |||||
QY 361 CACTCACTGCTG-----ATGAAATTTCTGAAAGCCCTGTGAAGAACTTC 405
   |||||
DB 373 TCAGATTCCTCGTGTGAGCGGTAAACAGAGATTAACCTTCAGAGCAAGTTCTAATATA 432
   |||||
QY 406 GAATCTGCTGATCATGTCAGTTGCTGTCACAGCTGCTGCTTATGTTGTCATCAG 465
   |||||
DB 433 GAAGGCTGCTGATCTTCCCTATCTCACTAAATATCTGAGGCTGTTATGTTGTAACAC 492
   |||||
QY 466 TTGGGCTCAAAATGGAATGGAACCATCTACCCCGCAACATATCATCTAATGATT 525
   |||||
DB 493 CTAAAAATGATCTATCATGGAAGTCAACAGCTCAAGAGGATGAGCTGTTATCTT 552
   |||||
QY 526 TGGGGTGTGCTACAGCAGTGGGTCAACAACTAATCCAAAGTTGCCAAATATCAATGCT 585
   |||||
DB 553 TGGGGTGTGCAACAGCAATTTGGAACAATGCTTATCAGCTGACCAACAGTTGATGCA 612
   |||||
QY 586 TATCTAAGATTTGAATCTGTTGCTTAAAGCATGAAAGCTTTTAAAGCTTATGCT 645
   |||||
DB 613 TTCACCAAAATTAATTAACCTTGATCCCTTAAACATGAACCTTCATGTAAGAAATACGCT 672
   |||||
QY 646 GCTGATGATGCTTGTGATCATGATGAGGCTTATGAGAGATCAATCGAAGTAT 705
   |||||
DB 673 GCTGATCATTTGTTGATTAACAAGATCTGAGATGCGAACAAGATTAAGAAAGGATAC 732
   |||||
QY 706 CCAAACTGCAACATGTTATGACGCTGTGGAAGAGCAATAGTATCCCGAGGCTTAT 765
   |||||
DB 733 CCAGAGAGCAAGTGTGTTGATGATGCTGAGAGTCTGGAACCTATATCTCAAAATTTAT 792
   |||||
QY 766 AAAAGTCAACGAGATGATGCTACCTGCAACATTTAAGAGTGTTCATGACCATGAA 825
   |||||
DB 793 AAGGTGACCGCTGATCACTAAC-----CATCAGTATTAATCAATTTGATTTAGAT 846
   |||||
QY 826 AGCATCTCGAAGAAATGAGAAAGATTAAGTAAATGATTAATCTTGTGTTATGCT 885
   |||||
DB 847 GACAAAGAGCTGGAAGAGCAAAATCAATTTGAGACAGAAAGTACATGTTGTTACTTA 906
   |||||
QY 886 GCATCTGTCAGAAATTTCTATTTGGGTGCAACAAGATTTCTGCTAGTCCAGAAATATCAT 945
   |||||

```

```

DB 907 ATGACTGTTTGAATATACATATGATATCCCTCCACTTACCAAGATTTGGAGTATAGA 966
QY 946 GAAGCAGAGTTAAATTCGTTAGTTTAAATCCACACCTTAAACAAGGTATTCAT 1005
   |||||
DB 967 AAAAAGGCATGATTTATCAATAGGGGTACACAA-AGATGATGACAGGTGACATTTAT 1025
   |||||
QY 1006 CATATGAATATTAAGTTTTCAGCAACGCTTAGATGATGTCACGCTCT 1055
   |||||
DB 1026 CACCTGCCAGTAATTT-----CAATAGCTTATGATCAATCCMAGCTCT 1071
   |||||

RESULT 6
CNS06K4B/c 942 bp DNA linear STS 10-JAN-2001
LOCUS T7 end of clone AT0A001G11 of library AT0A from strain CBS 4311
DEFINITION of Saccharomyces servazzii, sequence tagged site.
ACCESSION AL402417.1 GI:12161360
VERSION AL402417.1
KEYWORDS STS.
SOURCE Saccharomyces servazzii.
ORGANISM Saccharomyces servazzii.
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE 1 (bases 1 to 942)
Soucier,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
de-Montigny,J., Dujon,B., Durenne,P., Lepingle,A., Llorente,B.,
Malpertuy,A., Neveuglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
Saurin,W., Tekala,F., Toffano-Nicohe,C., Wesolowski-Louvel,M.,
Wincker,P. and Weissenbach,J.
Genomic exploration of the hemiascomycetous yeasts: 1. A set of
Yeast species for molecular evolution studies
FEBS Lett. 487 (1), 3-12 (2000)
20584711
JOURNAL MEDLINE
PUBMED 11152876
TITLE 2 (bases 1 to 942)
Casaregola,S., Lepingle,A., Bon,E., Neveuglise,C., Nguyen,H.,
Artiguenave,F., Wincker,P. and Galliardin,C.
Genomic exploration of the hemiascomycetous yeasts: 7.
Saccharomyces servazzii
FEBS Lett. 487 (1), 47-51 (2000)
20584717
JOURNAL MEDLINE
PUBMED 11152882
REFERENCE 3 (bases 1 to 942)
Genoscope.
Direct Submission
Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,
2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
segre@genoscope.cns.fr - Web : www.genoscope.cns.fr)
This STS is part of a random genomic sequencing program of thirteen
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
angusta, Debaryomyces hansenii var. hansenii, Pichia sorbicola,
Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
5 kb were prepared and both extremities were sequenced. See
keywords for description of this sequence and for the sequence of
the other extremity of this insert.
Location/Qualifiers
1..942
/organism="Saccharomyces servazzii"
/strain="CBS 4311"
/db_xref="taxon:27293"
/clone="AT0A001G11"
/clone_1b="AT0A"
/note="end : 17"
complement(320..>935)
/note="similar to Saccharomyces cerevisiae ORF YCR102c [
similarity to C.carbonum toxo gene ]
1 putative frameshift(s)
similar to Saccharomyces cerevisiae ORF YLR460c [
similarity to C.carbonum toxo protein ]"
misc_feature

```

misc_feature /evidence=not experimental
complement (<329..>920)
/note="similar to Saccharomyces cerevisiae ORF YNL134c (similarity to C.carbonum tox d gene)
1 putative frameshift(s)"
/evidence=not experimental

BASE COUNT 310 a 161 c 149 g 309 t 13 others

Query Match 9.3%; Score 103.6; DB 11; Length 942;
Best Local Similarity 49.5%; Pred. No. 1.1e-12;
Matches 303; Conservative 4; Mismatches 301; Indels 4; Gaps 2;

Y 502 CAACATATCTATCCATTATGATTTGGGGTGGTCTACACAGTGGCTC-AACACTAAT 560
DB 935 CAATTTAAAGGCAACATCBTGTGGGGTGGTCTACTCTTATCTCAGACATTTT 876
Y 561 CCAAGTGGCCAAACATATCAATGCTTATAGATTGTAACTGGCTCTAATAAGCA 620
DB 875 GCAACTATTACAAAAGTTTAACTGCCGAGAAAGTGTGTGCTCTAAGAAACA 816
Y 621 TGAAGAGCTTTTAAAGCTTATGCTGTGATGATGCTTATGATCATGATGACGGCT 680
DB 815 TGAAGCAATATATGAACTCGGTGACGAGAGTGTGATATCTGATGATGATG 756
Y 681 TATTGACGATCAATATGAGATTCCTGCAACATGTTATGACCTGTGGAG 740
DB 755 AATTTGATCAATTAAGCAAGTACCTGATTTGTTGTTATGATGATGCTGCTCAC 696
Y 741 CGAAGATGATCCCGAGGCTTAAAGTACAGCAGATGATCTACCTGCCATTAAT 800
DB 695 TCTAGACATTTTCAACCAAGTGTACATGCTGCCATGATGTTAATCACTGTCAT 636
Y 801 AGAAGTGTTCATATGACATGATGAAAGCTTCTGAAAGTCAAGAAAGATA---ATG 857
DB 635 CAATTATATGGGTTACTAGATGCTATTAACCAAGATTAAGCAACAGAAAGATG 576
Y 858 TAAATATGATATTAATCTTTGTTATGCTGATCTGCTCAAGAAATCTATGGGTGAC 917
DB 575 CTCTCTGAGGACATGCTCTTATACCTGTTTAAAGTTTGAATGATCACTTAAATGAATA 516
Y 918 AAGATTTCTGATGTCAGATATATCATGAAAGCAAGTAAATGTTAAGTTTAA 977
DB 515 CACAGTCAAGCTGAGNCCAGANACAGACAGNANCAITBAANATGTAATTAATTA 456
Y 978 TCCACACCTTAAACAGCGTATATCCATATGATTAATTAAGTTTTCAGCAACGCTT 1037
DB 455 TGAATAATTGACCAATGTTTCAATGAACATATCCCAATTAAGTGAACAAATGGGTT 396
Y 1038 AGATGATGTCGACGCTCTCACTGAGATATTAAGAAAGTAAACAAATAATTTAAGTA 1097
DB 395 AGATCTTACCATTCATTAATTAATGATTTGAAATGTTAAACCTGTGTAATAAT 336
Y 1098 TGTGGCAGGTT 1109
DB 335 TGTGGCAGCTKT 324

RESULT 7
LOCUS YSCDC91A/c 5355 bp DNA linear PLN 08-APR-1994
DEFINITION Saccharomyces cerevisiae cdc91 gene, complete cds.
ACCESSION U31649
VERSION U31649.1 GI:469465
KEYWORDS
SOURCE Saccharomyces cerevisiae DNA.
ORGANISM Saccharomyces cerevisiae
Bukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE
AUTHORS 1 (bases 1 to 5355)
TITLE Bi, E. and Pringle, J.R.
JOURNAL Sequencing and characterization of CDC91
Unpublished (1994)

FEATURES
source Location/Qualifiers
1..5355
/organism="Saccharomyces cerevisiae"
/db xref="taxon:4932"
gene 2540..3724
/gene="cdc91"
2540..3724
/gene="cdc91"
/codon start=1
/protein id="AAA34487.1"
/db xref="GI:469465"
/translation="MDSTRKVALGICIAIRLAVNSLPISQOOLDSVSEFSPYTSFR
SIQEGTYLIRNNIQQYRNGVNHPPILIFLSEFNSDLISITVLLIGLAVQITLV
TFAFNKLLKWLPGLLVAVPDLTLLSISRSIIPTNPAISSSYLCILAEQVLLSS
VMSISGLSVYPIILLPLGLMLSMRQRIASVLSILSLILFESYISQSSWS
PLTOYGGIIIEKEFPMGLMWYFPIEMDFPIFPAVENIPIAVITPEPTLYHK
OPEVAFICIGMIVLTQPYSPGADGFFSPFPPTPLGTYRLYPIISALPLHAIIVL
APTFHLMWVLGSSNSNFPYASLVYALAIASILVDLMMMLRIEVDGCIPIPKLKV
QT"

BASE COUNT 1621 a 975 c 933 g 1826 t

Query Match 9.3%; Score 103.2; DB 8; Length 5355;
Best Local Similarity 54.1%; Pred. No. 1.1e-12;
Matches 210; Conservative 0; Mismatches 178; Indels 0; Gaps 0;

Y 725 TTGACGCTGTGGAAAGCAGATAGATATCCCGAGGCTTAAAGTCAACGAGATATGTC 784
DB 5354 TCGACTGTGTGCGAATCAAGATACGCTTCAACAGTGTACAAATGTGCGCGATTAAC 5295
Y 785 TACCTGCCACATTTAAGATGCTTCCATGATCAATGAAAGCATCTCTGAAGAAATCA 844
DB 5294 AGATGCTACAAATGTTGATTAATAAAATTTGACAGAAAGAAACGTCAAAAAGGAAACA 5235
Y 845 GAAAGATATGTTAAATGATATTAATCTTGTGATGCTGATGCTGATGCAAGAAATTC 904
DB 5234 GAGACAAAGCTTACTATGATTAATTAAGCTATATTAATGATGATGATGATGATG 5175
Y 905 TATGGGTGCAACAAAGATTTCTGTGATGTCAGAAATATCATGAAAGCAGTTAAATTCG 964
DB 5174 CATTGGAAACATTTACTTTACAGCCGACCTGAAAGTGAAGACCTGCATTAATAATTA 5115
Y 965 TTAAGTTTAAATCCACCTTAACAGGATGATTCATATATGATTAATTAAGTTT 1024
DB 5114 TCAATATCAATCCAAAGATTAATGATGACAAATTCGCCATATTCAGTAAGGCTT 5055
Y 1025 TCAGCAACGCTTAAAGTATGTCCTCAGCTCTCACTGAGATTAATAAGGTAATAACA 1084
DB 5054 ATTAAGACGGGCTTTGTATGTTCTCATATCTTAAGACATCAAAATATGTTAAGACT 4995
Y 1085 AAAATGTTAATGATGTTGCGAGTTATA 1112
DB 4994 CTGTGAAACCTCGTGGCGATTAATA 4967

RESULT 8
LOCUS CETOXD 1506 bp DNA linear PLN 26-OCT-1995
DEFINITION C.carbonum tox d gene.
ACCESSION X92391
VERSION X92391.1 GI:1027510
KEYWORDS
SOURCE tox d gene.
ORGANISM Cochliobolus carbonum.
Bukaryota; Fungi; Ascomycota; Pezizomycotina; Dothideomycetes;
Pleosporales; Pleosporaceae; Cochliobolus.
REFERENCE
AUTHORS 1 (bases 1 to 1506)
TITLE Direct Submission
JOURNAL Submitted (09-OCT-1995) Y. Cheng, Michigan State University,
MSU-DOE Plant Research Laboratory, Room 210 Plant Biology Building,
East Lansing, Michigan 48824-1312, USA
REFERENCE 2 (bases 1 to 1506)

CDS
/gene="At2g44530"
/join(4242..4337,4472..4648,4876..5124,5395..5538,
5646..5813,5885..5954,6035..6116,6206..6380)
/gene="At2g44530"
/codon_start=1
/product="putative phosphoribosyl pyrophosphate
synthetase"
/protein_id="AAC27455.1"
/db_xref="GI:3341673"
/translation="MASIVQSPPTPALNRRSLIRPSSVRFPLKCNADPYKFDG
GNSGPHLITGDVPAFSRTRLEDSIYONTRLRIPSGANPLAQEISCVGLDG
KIKIRPADGEIVLOEISVRCGDVLTNOPTCPANDEIMELIMIDACRAKKT
AVIPRYGARAADKRTQGRBSIAKLVANLTGSGADRYLACDLSGSKGEPYVDH
VYQGVYLDYLAASKAISSEDLVVSVDVGVAARPAKLSAPLAIVDRKHANV
AEVWMLIGDVKGVAIMVDMIDITAGTISKGAALHQEGAREVYACTTAAVSPPAIS
RLSSGLFOEVITINNTIPLSEKNYFPQLTIVSANLIGETIWRVHDCSVSKT"
complement(6742..6763)
/rpt_family="AT_rich"
complement(6851..8620)
/gene="At2g44540"
/note="F16B22.3"
complement(join(6851..7168,7280..7561,7648..8301,
8399..>8620))
/gene="At2g44540"
complement(join(6851..7168,7280..7561,7648..8301,
8399..8620))
/codon_start=1
/product="putative cellulase"
/protein_id="AAC27456.1"
/db_xref="GI:3341674"
/translation="MSQSKNGSSQCLMTSICIVLVMSNARGAVSTNVAELQSLLY
FEAQRSGKLPNQKVTWRGDSALKDSDAHIDLTGYVYADGMMKGFPLATTTMLA
WSNIEGSESLRAHEQNSLRALKMATDYLIKHPQPNVLYGVGEGNSDKCMRPE
DMTPRTSYRIDQHPGSDLAGETAAAMAASIAFAPSDAVANILIGAKDLFAK
AHRGLYONSI PNAGGVYASGYEDELMAAAMLRATNDQIYDLYTEATGSPRVF
AMDKFGAQLVAKLAEKGVESSEIYKYSMAEOPICNCAQKSNVKKTPGGL
LYFLPMNLQYTTAAFTVLSAYSKYLTAASIKIDCPGALAQSDLLQVARSQVYILG
NPKMSYMGVGTNYPKPHRAASIVISITDKTPTVCSGGLYKNNRPNPNVLAG
AVVGGPDNDVYGDERSNFOQAEPAFTVTAFLVGLALVF"
complement(12424..14164)
/gene="At2g44560"
/note="F16B22.5"
complement(join(12424..12741,12834..13115,13199..13852,
13943..>14164))
/gene="At2g44560"
complement(join(12424..12741,12834..13115,13199..13852,
13943..14164))
/gene="At2g44560"
/codon_start=1
/product="putative glucanase"

gene
repeat_region
mRNA
CDS
/gene="At2g44540"
complement(join(6851..7168,7280..7561,7648..8301,
8399..>8620))
/gene="At2g44540"
complement(join(6851..7168,7280..7561,7648..8301,
8399..8620))
/codon_start=1
/product="putative cellulase"
/protein_id="AAC27456.1"
/db_xref="GI:3341674"
/translation="MSQSKNGSSQCLMTSICIVLVMSNARGAVSTNVAELQSLLY
FEAQRSGKLPNQKVTWRGDSALKDSDAHIDLTGYVYADGMMKGFPLATTTMLA
WSNIEGSESLRAHEQNSLRALKMATDYLIKHPQPNVLYGVGEGNSDKCMRPE
DMTPRTSYRIDQHPGSDLAGETAAAMAASIAFAPSDAVANILIGAKDLFAK
AHRGLYONSI PNAGGVYASGYEDELMAAAMLRATNDQIYDLYTEATGSPRVF
AMDKFGAQLVAKLAEKGVESSEIYKYSMAEOPICNCAQKSNVKKTPGGL
LYFLPMNLQYTTAAFTVLSAYSKYLTAASIKIDCPGALAQSDLLQVARSQVYILG
NPKMSYMGVGTNYPKPHRAASIVISITDKTPTVCSGGLYKNNRPNPNVLAG
AVVGGPDNDVYGDERSNFOQAEPAFTVTAFLVGLALVF"
complement(12424..14164)
/gene="At2g44560"
/note="F16B22.5"
complement(join(12424..12741,12834..13115,13199..13852,
13943..>14164))
/gene="At2g44560"
complement(join(12424..12741,12834..13115,13199..13852,
13943..14164))
/gene="At2g44560"
/codon_start=1
/product="putative glucanase"

Query Match 4.3%; Score 48.2; DB 8; Length 109375;
Best local Similarity 50.7%; Pred. No. 1.2; Mismatches 116; Conservative 0; Mismatches 113; Indels 0; Gaps 0;
Matches 116; Conservative 0; Mismatches 113; Indels 0; Gaps 0;
Query 800 TAGAAGTGTTCGATGACCATTAAGGATTCCTGAGAAATCAAGAAATTAATGTTA 859
|||||
Db 19728 TAGTATATGAAACAACACTAATTTGTCGACATTATTCGTTGCTGGAATTTGTAAGTTA 19787
|||||
Qy 860 AATGATATTAATCTTTGTTGATATCGGATCGGCAAGAAATTCATTTGGTGCAACA 919
|||||
Db 19788 AAAATTAATTAATTTAATTTGTAAGGCTGTTTACATTAATAGTTGGTACACAA 19847
|||||
Qy 920 GATTTCTGCTAGTCAGAAATTCATGAAGCCAGGTTAAATTCGTTAAGTTATAAATC 979
|||||
Db 19848 TCTCTCCATGAACAATATAAACAATTAACAATTAACAAAAATGAATATAGTC 19907
|||||
Qy 980 CACACCTTAACAACGGTATATTCATCATATGAATATTTAAAGTTTTCAG 1028
|||||
Db 19908 ATTACTTTTCCAGAGTTTACTGTGCGATTAAACAATACGATGTCAG 19956
|||||
-RESULT 10
PFMAL1P2 98734 bp DNA linear HTG 11-AUG-1999
LOCUS Plasmodium falciparum 3D7 chromosome 1, *** SEQUENCING IN PROGRESS
DEFINITION *** in unordered pieces.
ACCESSION AL0311745
VERSION AL0311745.7 GI:5731893
/protein_id="AAC27458.1"
/db_xref="GI:3341676"
/translation="MSQKIGSSQCLMTSICIVLVMSNARGAVSNYGEALTKSLY
FAORSRGKLPNSORVTWRGDSALRDGSDAHVDLTGTYADGMMKGFPLATTTMLA
WSNIEGSESLRAHEQNSLRALKMATDYLIKHPQPNVLYGVGEGNSDKCMRPE
DMTPRTSYRIDQHPGSDLAGETAAAMAASIAFAPSDAVANILIGAKDLFAK
AHPGLYONSI PNAGGVYASGYEDELMAAAMLRATNDQIYDLYTQSGTGPRFA
FEMDKEPVAQQLVAKLAEKGVESNGKIAEYSMAEOPICNCAQKSNVKKTPGGL
LYFLPMNLQYTTAAFTVLSAYSKYLTAASIKIDCPGALAQSDLLQVARSQVYILG
SNPKMSYMGVGTNYPKPHRAASIVISITDKTPTVCSGGLYKNNRPNPNVLAG
GAVGPPNDNDVYGDERTDYOHAEBAPATPAAFVGLAAVA"
complement(14607..14630)
/rpt_family="AT_rich"
/rpt_family="AT_rich"
complement(14935..16778)
/gene="At2g44570"
/note="F16B22.6"
complement(join(14935..15252,15445..15729,15807..16460,
16557..>16778))
/gene="At2g44570"
complement(join(14935..15252,15445..15729,15807..16460,
16557..16778))
/codon_start=1
/product="putative glucanase"
/protein_id="AAC27459.1"
/db_xref="GI:3341677"
/translation="MSCISQCFITIKSICIVLVITNGAVSANSYGEALTKSLY
FEAQRSGKLPNDORVTWRGDSALRDGSDAHVDLTGTYADGMMKGFPLATTTMLA
WSNIEGSESLRAHEQNSLRALKMATDYLIKHPQPNVLYGVGEGNSDKCMRPE
DMTPRTSYRIDQHPGSDLAGETAAAMAASIAFAPSDAVANILIGAKDLFAK
AHPGLYONSI PNAGGVYASGYEDELMAAAMLRATNDQIYDLYTQSGTGPRFA
FEMDKEPVAQQLVAKLAEKGVESNGKIAEYSMAEOPICNCAQKSNVKKTPGGL
LYFLPMNLQYTTAAFTVLSAYSKYLTAASIKIDCPGALAQSDLLQVARSQVYILG
SNPKMSYMGVGTNYPKPHRAASIVISITDKTPTVCSGGLYKNNRPNPNVLAG
GAVGPPNDNDVYGDERTDYOHAEBAPATPAAFVGLAAVA"
complement(16815..16914)
/rpt_family="TA)n"
complement(17216..17237)
/rpt_family="AT_rich"
complement(18506..23922)
/gene="At2g44580"
/note="F16B22.7"

KEYWORDS HTG; HTGS PHASE1.
SOURCE Plasmodium falciparum 3D7.
ORGANISM Plasmodium falciparum 3D7.
REFERENCE Eukaryota; Alveolata; Apicomplexa; Hemosporida; Plasmodium.
AUTHORS Bowman, S., Churcher, C., Harris, B., Harris, D., Lawson, D., Quail, M., and Barrett, B.
TITLE Direct Submission
JOURNAL Submitted (24-SEP-1998) P. falciparum Genome Sequencing Consortium, The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK

COMMENT
On Aug 12, 1999 this sequence version replaced gi:5706497.
For more information about this sequence or the Malaria Project, see http://www.sanger.ac.uk/Projects/P_falciparum. IMPORTANT: This sequence is unfinished and does not necessarily represent the correct sequence. Work on the sequence is in progress and the release of this data is based on the understanding that the sequence may change as work continues. The sequence may be contaminated with foreign sequence from E.coli, yeast, vector, phage etc.
Order of segments is not known; 800 n's separate segments.
* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

FEATURES
SOURCE
1..98734
/organism="Plasmodium falciparum 3D7"
/strain="3D7"
/db_xref="taxon:36329"
/chromosome="1"
BASE COUNT 37227 a 8619 c 9384 g 35504 t 8000 others
ORIGIN

Query Match 4.3%; Score 47.6; DB 2; Length 98734;
Best Local Similarity 43.2%; Pred. No. 1.6;
Matches 227; Conservative 0; Mismatches 299; Indels 0; Gaps 0;

QY 563 AAGTGGCCAAACATTCATGCTTATCTACCAAGTGTACTGTCTTAAAGCATG 622
Db 85347 AATGATGAT 85406

QY 623 AAAAGCTTTAAAGCTTATGCTGATGATGTCTTTCATCATCATGATGACGCTTA 682
Db 85407 ATATATTTGAT 85466

QY 683 TTGACGACATCAATCGAAGTATCCAAACCTGCACATGTTTGAACGCTGGGAAGCG 742
Db 85467 GTGATGAT 85526

QY 743 AAGATAGTATCCCGAGGCTTAAAGTACACAGCATAGTCTACCTGCACATTTATG 802
Db 85527 GTCAT 85586

QY 803 AAGTGGTTCATGACCATTCGAAAGCATCTCGAAGAAATCGAAGAAATATGTTAAA 862
Db 85587 ATTAACAAAGAAAGTATGATGATGATGATGATGATGATGATGATGATGAT 85646

QY 863 TTGATATATCTTTGTTGATCGTGATCTGTCAAGAAATTTATTTGGTGCAACAGAT 922
Db 85647 TAGATATATGTTTATCTATCTGTTAGTATTAACAATATATGATGATATATAT 85706

QY 923 TTCCTGCTAGTCCAGATTCATGAAAGCAAGTTAAATTCGTTAAGTTAATCCAC 982
Db 85707 GTAGGATATATATATGATGATGATGATGATGATGATGATGATGATGAT 85766

QY 983 ACCCTTACACAGGTGATTCATCATGATATATTAAGTTTTCAGCAACGCTTAGATG 1042
Db 85767 ATGATATATATATATGATGATGATGATGATGATGATGATGATGATGAT 85826

QY 1043 ATGTCCACGCTCTCACTGAAGGTATTAAGAAAGTAAACAAAAA 1088
Db 85827 ATATATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 85872

RESULT 11
AE007561
LOCUS 13844 bp DNA linear BCT 27-JUL-2001
DEFINITION Clostridium acetobutylicum ATCC824 section 49 of 356 of the complete genome.
ACCESSION AE007561 AE001437
VERSION AE007561.1 GI:15023317
KEYWORDS
SOURCE
ORGANISM
Clostridium acetobutylicum.
Clostridium acetobutylicum
Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae; Clostridium.
1 (bases 1 to 13844)
Nolling, J., Breton, G., Omelchenko, M. V., Markarova, K. S., Zeng, Q., Gibson, R., Lee, H. M., Dubois, J., Qiu, D., Hitti, J., Wolf, Y. I., Tatusov, R. L., Sabathe, F., Doucette-Stamm, L., Soucaille, P., Daly, M. J., Bennett, G. N., Koonin, E. V. and Smith, D. R.
Genome sequence and comparative analysis of the solvent-producing bacterium Clostridium acetobutylicum
J. Bacteriol. 183 (16), 4823-4838 (2001)

REFERENCE
AUTHORS
MEDLINE
PUBMED
11466286
2 (bases 1 to 13844)
21359325
Childrens, D., Zeng, Q. and Smith, D. R.
Direct Submission
Submitted (24-JUL-2001) GTC Sequencing Center Production, Finishing, and Bioinformatics teams, Genome Therapeutics Corp., 100 Beaver Street, Waltham, MA 02453-8443, USA

FEATURES
SOURCE
1..13844
/organism="Clostridium acetobutylicum"
/strain="ATCC 824"
/db_xref="ATCC:824"
/db_xref="taxon:1488"
135..1466
/gene="CAC0460"
135..1466
/gene="CAC0460"
/transl_table=1
/codon_start=1
/product="CBS-domain containing protein, YHP B. subtilis ortholog"
/protein_id="AAK78440.1"
/db_xref="GI:15023318"
/translation="MEADPPDDNNILQFVLAVTLVNAFSAEMSIVSLNTRMKI
LSEGNRAQILEKLIKPEFELSTVQGITLAFSSAPATGSLDMFGYLLKINI
PYSSOVALIVTVLVSFTLVGELFPRIALOSEKIAMLAAPILVSKITVPEVK
LLASATVLAIRITGMNNKLEKKSIEIKSIVAGOGVGVINSEKEMDISPEFD
RIAEVMTPRREIVLINIDTPLNKTVDELEIKSRVYFPGDSNITIGLIXKDFIV
EARKVGFENVIKNIHSYFVPERKINIEFLKLAQAKKMAVLAIDYGFSGIVTI
EDLIEVWGNISDEYEDDNYIKIDESTFIKQIIPRIENBOIDLDELDETDYDT
LGGLILKILIGHIPEDNEQSIKYKGIIFEIKVEKEKRIEKEVHLQKOLE"
1493..1891
/gene="CAC0461"
1493..1891
/gene="CAC0461"
/transl_table=1
/codon_start=1
/product="Mercuric resistance operon regulatory protein, Mer family"
/protein_id="AAK78441.1"
/db_xref="GI:15023319"
/translation="WRYKIGELSKIVNLSEITRYYEKRLIPKPERFNNRYKYS
MYIRLKIIRAKYVGFILHISNPLSEVSTSIEDIDLSSTIMNKYQIQIEKCERLN
YKQOLKFEIEINDLOCPVLRILINEKNT"
1955..3151
/gene="CAC0462"
1955..3151
/gene="CAC0462"
/codon_start=1
/transl_table=1

/product="Protein of short-chain alcohol dehydrogenase family"
/protein_id="AAK78442.1"
/db_xref="GI:15023320"
/translation="MIYKAKFVGFIRDVHPYGCRRBVNLQIDYCKAIGFRPKVU
IVASGFGIATRISVAFPGPEAHITGVSENGATDRICAGWNNINPEKFAKKG
LVAKNITEDAFSNETKDKYKIKDEFGKIDLVYSLAARRKDYTGAVYSRITTI
LGDPEPTIDVERDEITLKKVSASIEIEETKRWGDEWQCELELYEDCFDPA
TTIAYSIGSPRTYKIRREGTIGIAKDLDDAKLNEKLNRIYIGARAFVSKALVT
KASAYIPTPELYAAILYKWKERKNIHENCIMQIERMSEKIKYENKIOFDDKRLMD
DLERKQVDEVDRIINMTIPENFKEISDYKKEPMNLNGFDLDGVDSKDLIEL
LRKLEP"
/gene="CAC0463"
/complement (3180. .4226)
/gene="CAC0463"
/codon_start=1
/transl_table=11
/product="Serine protease Do (heat-shock protein)"
/protein_id="AAK78443.1"
/db_xref="GI:15023321"
/translation="MKTHKLISIVLASVLLTSFTGCTFANKKKTIPTDANQMAST
SVSPITKASKVLPSVGVTTIYIERDNTKEGVSGMTIDSSGYILTNNHAGTGS
KDLKSLVQSSSIGAKPLMANESLDSIIRKIDKNLQAVTLGDSGVLDGETALIG
PLGAFORTYTSIGISVANRTVAGBSVFEEDLQTDASINPNSGSPILIDANGNIG
VNSAKITSAIGIFAVPINIVKPLSLKLTGQPKTPIVIGITGLDKSQNGYLANTEK
GIYVNI SPNSGAAAGINKGDIILSVNGKNITMELRESIYITIGANNVLSLKTGA
SGEKTIVVKIQSST"
/gene="CAC0464"
/gene="CAC0464"
/codon_start=1
/transl_table=11
/product="Protein, related to divalent cations, tolerance
Cura family"
/protein_id="AAK78444.1"
/db_xref="GI:15023322"
/translation="MQINQPKIEIFIPKEVEKLGKLSANVKGKIGNYDHCNATIV
EGYWRPLEGSEPEGEIGKVCHECKIEMKCKREVENALKIRVHPYEPEVINVI
PLNQLFENNR"
/complement (4841. .5632)
/gene="CAC0465"
/complement (4841. .5632)
/gene="CAC0465"
/codon_start=1
/transl_table=11
/product="Predicted transcriptional regulator,
dica/hib/janser family"
/protein_id="AAK78445.1"
/db_xref="GI:15023323"
/translation="MRDSNLEIKILNKKIYDETSIFIHNLKKLQLEGTOKKLASKI
GVSEDLISKYSGDAPPALETILYICKLVITDEFLNIPLTSSYFEDMYLDGTGED
IFKODYCEYFVNLSGKSGSHESNLRITKNCIFPIYIKGEVTKFKGTYKSDLV
FNLVSRREYGAITIMIKPSLNKSKYIGGIAMLSDANSKRCCKIILSSKINRE
LSFDILANFLCFRSEKSGFONIKISRYEDKATDPTTNLEETKL"
/gene="CAC0466"
/gene="CAC0466"
/complement (3180. .4226)
/gene="CAC0466"
/codon_start=1
/transl_table=11
/product="Hypothetical protein"
/protein_id="AAK78446.1"
/db_xref="GI:15023324"
/translation="NMLRKVIYDIKSLCEYEFOLKITYPDKIYGYIYKNSIEDK
YIEPMILIDEPTEGELINLIKTIHDKLKNSKRYGVSJLDGVKIDLMKPIIYVEN
GKAKGVMANIDYFTMWLVKNKAVGKSPSIDSLKGEF"
/gene="CAC0467"
/gene="CAC0467"
/complement (3180. .4226)
/gene="CAC0467"
/codon_start=1

/transl_table=11
/product="Unclassified membrane protein, homolog of
YAH B, subfamily 1"
/protein_id="AAK78447.1"
/db_xref="GI:15023325"
/translation="NNIOILLINFTTIFIGTLAYSRILVGVTKGIAISFTVNI
LTVSRATVTPPEPLITNVAEKLYSNLIYFNSIILVSGAASILGAVLITPORMF
SGVAFPSLEKSIPLRVHSMTRKRIYSKQCVLPSRSKVEKFNKLPKRIYVNF
ISVALLVVALAPPIYAFKIPMLRATCVLSSVNGIAQIMLTIFIDPQMSIMTEVM
EKKECEDEPKNCVLAMVSGKVGTPAALFILIPASFAIVAKGVDFISRI"
/gene="CAC0468"
/gene="CAC0468"
/complement (3180. .4226)
/gene="CAC0468"
/codon_start=1
/transl_table=11
/product="HAD superfamily hydrolase"
/protein_id="AAK78448.1"
/db_xref="GI:15023326"
/translation="MIKALMDSGVLNKPATGENFIPKFFYYVDWERFTIPLIKR
KSAFPAKANKYIKKQKLTITEBEYKCFLEYKILSKELYQKLKNNDLIEITDELVN
YKRYDPEYKDAIYEIPLSKYKLAIVISAMPLEGVFKKAGRTFSSVSVGVVT
KENEMVYKTAISELRVLPEEAVFIDNIRNCEGAVKLGKSPVLSRSKRIYANNKLFN
RRIKSIRNKKOLEIV"
/gene="CAC0469"
/gene="CAC0469"
/note="spma"
/codon_start=1
/transl_table=11
/product="Spore maturation protein A (gene spma)"
/protein_id="AAK78449.1"
/db_xref="GI:15023327"
/translation="MINYIWFIFIVGTIVGVLSNGSTLSSSIINSTEVSVMKLSL
IGIMCLMGVMRIARSGITDKLSIVLPILKIVKDIKNNWNNWPMNTSMNG
LGNATPREGIKTMEKQKINRKSRATMDKAEFLVYINACIOLEPTTIIISIAACGS
NPAITVPIVTSVTVAINGLIYCKILDRHF"
/gene="CAC0470"
/gene="CAC0470"
/note="spmb"
/codon_start=1
/transl_table=11
/product="Spore maturation protein B (gene spmb)"
/protein_id="AAK78450.1"
/db_xref="GI:15023328"
/translation="NNYISSALIPITISIVLYGMIKKVKYCEVGAQKDIRICVN
IPPYILAMLSAIFRSGAMQYLINILKPVKIKITGPSEVLPILIVKPLSSGALGV
FAEILIKRGADTYIGILASILMGSTETTFYITVYGAIKIRNIRHTTMAALMDWTS
IIVALIVRFEKI"
/gene="CAC0471"
/gene="CAC0471"
/complement (3180. .4226)
/gene="CAC0471"
/codon_start=1
/transl_table=11

Query Match
Best Local Similarity 5.3%; Score 47.4; DB 1; Length 13844;
Matches 102; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

QY	809	TTCCATGACCATGTAAGCATTCCTGAGAAATCAGAAAGATATGTTAAATTGATA	868
DB	2163	TTGAGATACCTATGAAACAGAGCTACAGATAGAAATAGAAACAGCGGATGATATA	2222
QY	869	TTACTTTGTTGATCGTCATCTGGTCAAGAAATTCATTTGGGTGCAACAATTTCCG	928
DB	2223	ATPACATATTTTAAAGATTTGCTAAAAAAGAGATTAGTTGCAAAAACTTCATTG	2282
QY	929	CTAGTCCAGATATATATGAGCCACAGCTAAATTCGTTAAGTTTAAATTCACACCTTA	988

Db 2283 AGAGATCCCTTCTATGATAACCAAGATTAAGTATTAAGTATTAAGATGATTTG 2342

QY 989 ACAACGGATAT 1001

Db 2343 GTAATAATGATTT 2355

RESULT 12

LOCUS HSJ106901

DEFINITION HSJ106901 92357 bp DNA linear PRI 04-APR-2001

Human DNA sequence from clone RPS-106901 on chromosome 20. Contains part of the gene for a novel protein similar to a low density lipoprotein-related protein LRP16 and STSs, complete sequence.

ACCESSION AL049633

VERSION AL049633.3 GI:4775651

KEYWORDS HTG; LRP16.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

AUTHORS 1 (bases 1 to 92357)

TITLE Direct Submission

JOURNAL Submitted (15-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk

COMMENT On May 11, 1999 this sequence version replaced gi:4678551. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WORMBEP; Information on the WORMBEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormep/ This sequence is the entire insert of clone RPS-106901. The true left end of clone RPS-106901 is at 92256 in this sequence. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. This sequence was generated from part of bacterial clone contigs of human chromosome 20, constructed by the Sanger Centre Chromosome 20 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr20> RPS-106901 is from the library RPCI-5 constructed by the group of Pletier de Jong. For further details see <http://www.chori.org/bacpac/home.htm> VECTOR: pCYPAC2.

repeat_region /note="LTR37B repeat: matches 1. .468 of consensus" 5707. .5787

repeat_region /note="MIR repeat: matches 66. .154 of consensus" 6118. 6221

repeat_region /note="LMCS repeat: matches 7312. .7409 of consensus" 6401. .6736

repeat_region /note="LMCS repeat: matches 7560. .7913 of consensus" 6803. .6850

repeat_region /note="24 copies 2 mer tt 75% conserved" 6853. 7163

repeat_region /note="Alusq repeat: matches 1. .312 of consensus" 7333. .7704

repeat_region /note="THEIC repeat: matches 1. .371 of consensus" 7705. .9306

repeat_region /note="THEIC-internal repeat: matches 1. .1580 of consensus" 9307. .9684

repeat_region /note="THEIC repeat: matches 1. .371 of consensus" 9687. .9770

repeat_region /note="S repeat: matches 7. .90 of consensus" 10529. 10623

repeat_region /note="L2 repeat: matches 2642. .2749 of consensus" 10724. .10795

repeat_region /note="MIR repeat: matches 181. .252 of consensus" complement(11896. .12355)

repeat_region /note="match: STS: Em:HSPF12C11 Em:294565" 12016. 12114

repeat_region /note="MIR repeat: matches 36. .130 of consensus" 12777. 12823

repeat_region /note="MIR repeat: matches 77. .123 of consensus" 12904. .13054

repeat_region /note="MER5A repeat: matches 11. .189 of consensus" 14132. 14435

repeat_region /note="Alu repeat: matches 1. .304 of consensus" 15118. 15242

repeat_region /note="L2 repeat: matches 2125. .2253 of consensus" 15257. 15299

repeat_region /note="MER5B repeat: matches 7. .49 of consensus" 15300. 15351

repeat_region /note="L3 copies 4 mer ttg 82% conserved" 15352. 15460

repeat_region /note="MER5B repeat: matches 44. .147 of consensus" 15460. 15919

repeat_region /note="L2 repeat: matches 2275. .2709 of consensus" 16557. 16805

repeat_region /note="MIR repeat: matches 4. .262 of consensus" 17067. 17516

repeat_region /note="LTR2B repeat: matches 1. .444 of consensus" 18905. 18964

repeat_region /note="L2 repeat: matches 2689. .2748 of consensus" 18914. 18999

repeat_region /note="MIR repeat: matches 175. .257 of consensus" 18990. 19101

repeat_region /note="L2 repeat: matches 2598. .2743 of consensus" 19107. 19464

repeat_region /note="LTR1A1 repeat: matches 1. .363 of consensus" 19482. 19561

repeat_region /note="MER69 repeat: matches 2426. .2505 of consensus" 19985. 20378

repeat_region /note="LM2 repeat: matches 5686. .6121 of consensus" 20395. 20462

repeat_region /note="MER69 repeat: matches 1. .68 of consensus" 21736. 21783

repeat_region /note="12 copies 4 mer tcat 77% conserved" 21899. 22178

repeat_region /note="AluSg1 repeat: matches 7. .283 of consensus" 22507. 22849

repeat_region /note="MER61B repeat: matches 7. .425 of consensus" 23004. 23082

repeat_region /note="L2 repeat: matches 2634. .2709 of consensus" 23075. 23372

repeat_region /note="L2 repeat: matches 2408. .2697 of consensus" 23680. 24013

repeat_region	/note="L2 repeat: matches 2038. .2384 of consensus"
repeat_region	/note="L2 repeat: matches 515. .617 of consensus"
repeat_region	/note="L2 repeat: matches 26. .221 of consensus"
repeat_region	/note="L2 repeat: matches 64. .168 of consensus"
repeat_region	/note="L2 repeat: matches 2654. .28647 of consensus"
repeat_region	/note="L2 repeat: matches 4167. .6304 of consensus"
repeat_region	/note="L2 repeat: matches 29755. .30565 of consensus"
repeat_region	/note="L2 repeat: matches 5473. .6300 of consensus"
repeat_region	/note="L2 repeat: matches 1. .468 of consensus"
repeat_region	/note="L2 repeat: matches 7. .1228 of consensus"
repeat_region	/note="L2 repeat: matches 3105. .33174 of consensus"
repeat_region	/note="L2 repeat: matches 2671. .2737 of consensus"
repeat_region	/note="L2 repeat: matches 35600. .35623 of consensus"
repeat_region	/note="L2 repeat: matches 35695. .35860 of consensus"
repeat_region	/note="L2 repeat: matches 23. .186 of consensus"
repeat_region	/note="L2 repeat: matches 35966. .36119 of consensus"
repeat_region	/note="L2 repeat: matches 36. .198 of consensus"
repeat_region	/note="L2 repeat: matches 37193. .37331 of consensus"
repeat_region	/note="L2 repeat: matches 81. .224 of consensus"
repeat_region	/note="L2 repeat: matches 37805. .37859 of consensus"
repeat_region	/note="L2 repeat: matches 49. .115 of consensus"
repeat_region	/note="L2 repeat: matches 37957. .38255 of consensus"
repeat_region	/note="L2 repeat: matches 72. .410 of consensus"
repeat_region	/note="L2 repeat: matches 39627. .39920 of consensus"
repeat_region	/note="L2 repeat: matches 2362. .2707 of consensus"
repeat_region	/note="L2 repeat: matches 41385. .41697 of consensus"
repeat_region	/note="L2 repeat: matches 41854. .42272 of consensus"
repeat_region	/note="L2 repeat: matches 1377. .1805 of consensus"
repeat_region	/note="L2 repeat: matches 42557. .42851 of consensus"
repeat_region	/note="L2 repeat: matches 43341. .43987 of consensus"
repeat_region	/note="L2 repeat: matches 5488. .6171 of consensus"
repeat_region	/note="L2 repeat: matches 44255. .44294 of consensus"
repeat_region	/note="L2 repeat: matches 44681. .45000 of consensus"
repeat_region	/note="L2 repeat: matches 29. .364 of consensus"
repeat_region	/note="L2 repeat: matches 46252. .46605 of consensus"
repeat_region	/note="L2 repeat: matches 46606. .46920 of consensus"
repeat_region	/note="L2 repeat: matches 46921. .46982 of consensus"
repeat_region	/note="L2 repeat: matches 988. .1050 of consensus"
repeat_region	/note="L2 repeat: matches 47017. .47040 of consensus"
repeat_region	/note="L2 repeat: matches 47321. .47738 of consensus"
repeat_region	/note="L2 repeat: matches 16. .448 of consensus"
repeat_region	/note="L2 repeat: matches 47944. .48244 of consensus"
repeat_region	/note="L2 repeat: matches 3552. .3871 of consensus"
repeat_region	/note="L2 repeat: matches 48261. .48573 of consensus"
repeat_region	/note="L2 repeat: matches 7218. .7522 of consensus"
repeat_region	/note="L2 repeat: matches 49230. .49257 of consensus"
repeat_region	/note="L2 repeat: matches 2626. .2653 of consensus"
repeat_region	/note="L2 repeat: matches 49339. .49430 of consensus"
repeat_region	/note="L2 repeat: matches 137. .227 of consensus"
repeat_region	/note="L2 repeat: matches 49466. .49525 of consensus"
repeat_region	/note="L2 repeat: matches 7856. .7913 of consensus"

Query Match	4.3%	Score 47.4;	DB 9;	Length 92357;
Best Local Similarity	52.8%	Pred. No. 1.8;		
Matches 102;	Conservative 0;	Mismatches 91;	Indels 0;	Gaps 0;

[illegible]

Db	6418	CTATGTAATAAATAATTACATTTTAAATAATGAATAATACTATATATATATCAAGAAAGAA	6477
Qy	856	GTTAAATATGATATTACTTTGTTGATCGTGCAATCTGGTCAAGAAATTTCAATTGGGTGCA	915
Db	6478	TTTAAATATTAATTTTCAGTGTGTATAGTATATATTTTAACTATTTCAATATTTTGA	6537
Qy	916	ACAAGATTTTCTG	928
Db	6538	AATATATATACGG	6550

Db. 6538 AATATATATACGG 6550

RESULT 13	AC008057	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM
	AC008057	119082 bp	DNA	linear	HTG 11-OCT-2000			
		Homo sapiens chromosome 20 clone PI-10263, WORKING DRAFT SEQUENCE,						
		3 unordered pieces.						
	AC008057							
	AC008057.4	GI:10765024						
	HTG: HTGS_PHASE1; HTGS_DRAFT.							
	Homo sapiens.							
	Homo sapiens							

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

1 (bases 1 to 115082)
McCombie, W. R.
Human Genomic Sequence, Chromosome 20
Unpublished
2 (bases 1 to 115082)
McCombie, W. R.
Direct Submission
Submitted (17-JUL-1995) Lita Annenberg Hazen Genome Sequencing

COMMENT

COMMENT

On Oct 11, 2000 this sequence version replaced g4:9954568.

* NOTE: This is a 'working draft' sequence. It currently

* consists of 3 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

*	1	68531: contig of 68531 bp in length
*	68532	68631: gap of unknown length
*	68632	contig of 33932 bp in length
*	102564	102663: gap of unknown length
*	102564	119082: contig of 16419 bp in length.

FEATURES	Location/Qualifiers
source	1. .119082 *

	a	c	g	t	others
BASE COUNT	37455	22465	21882	36187	1093
ORIGIN					

Query Match	4.3%	Score 47.4	DB 2	length 119082
Best Local Similarity	52.8%	Pred. NO. 1.8		
Matches 102; Conservative	0	Mismatches 91	Indels 0	Gaps 0

QY 736 GGAAGCGAATAGTATCCCGAGGCGCTATAAAGTCACAGCAGATAGTCTAACCTGCCACA 795

QY 796 TTATTGAAAGTGTTCCAATGACCATTTGAAGCATTCCTGAAGAAATCAGAAAGAATAT 855

Db 12769 .CATGTAAAAAAGTTCAATTTAAAAATGAATATATACTTATTATATACAGGAAAAAGAA 1282

Qy 856 GTTAAATTTGATTTACTTTGTTGTAATCGTCGACTGTCAGAAATTCATATGGGTCGA 915

b 12829 TTTAAATATTTAATTTTCAGTGTGTAATAGATTAATTTTAACTAATTTCTAATATTTTAGA 12888

Db 12829 TTTAATATTTCAGTGTGTATTAGTATTATTTTAACTATTCATATTTTGA 1288

QY 916 ACAAGATTCTCG 928
 DB 12889 AATATATATACCG 12901

RESULT 14
 AC007630 169494 bp DNA linear PRI 22-NOV-2000
 LOCUS Homo sapiens chromosome 20 clone pl-7739, complete sequence.
 DEFINITION AC007630 GI:11245571
 VERSION AC007630.4 GI:11245571
 KEYWORDS HTG.
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE
 AUTHORS MCombie, W.R.
 TITLE Direct Submission
 JOURNAL Submitted (21-MAY-1999) Lita Annenberg Hazen Genome Sequencing Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, NY 11724, USA
 2 (bases 1 to 169494)
 de la Bastide, M., Gnoj, L., Dedhia, N.N., Matero, A., Ning Huang, E., O'Shaughnessy, A., Preston, R., Rodriguez, M., Schut, K., Shah, R., Shekher, M., Spiegel, L., Swaby, I., Vill, D. and MCombie, W.R.
 DIRECT SUBMISSION
 Submitted (02-JUL-1999) Lita Annenberg Hazen Genome Sequencing Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, NY 11724, USA
 3 (bases 1 to 169494)
 MCombie, W.Richard.
 DIRECT SUBMISSION
 Submitted (18-NOV-2000) Cold Spring Harbor Laboratory
 T10 transposon removed.
 4 (bases 1 to 169494)
 MCombie, W.Richard.
 DIRECT SUBMISSION
 Submitted (21-NOV-2000) Cold Spring Harbor Laboratory
 T10 transposon removed.
 On Nov 21, 2000 this sequence version replaced gi:5333384.
 Clone pl-7739 contains a 3 kb overlap with clone 48F5. The overlap ends at base 2940. From position 154930-155350 the subclones were single stranded and had single chemistry.

FEATURES
 source
 1..169494
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="20"
 /clone="pl-7739"
 134563
 /note="Site of Tn10 transposon insertion. Tn10 transposon, localing 1328bp, was removed from final sequence."
 /evidence=not experimental
 BASE COUNT 52335 a 31812 c 32399 g 52939 t 9 others
 ORIGIN

Query Match 4.3%; Score 47.4; DB 9; Length 169494;
 Best Local Similarity 52.8%; Pred. No. 1.7;
 Matches 102; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

QY 736 GGAAGGAGATGATGATCCCGAGGCTATTAAGTACAGAGATGCTACTGCGACA 795
 DB 12003 GAAAGCAGACAGAAAGCTTACGATTAAGAAAGGACAAAGATCTTAAATCAACAA 11944
 QY 796 TTATTAGAGTGGTCCCAATGACCATGGAAGCATTCGAGAGAAATCAGAAAAGATAT 855
 DB 11943 CTATGTAAGAAATTTCAATTTAAATGAATTAATTAATTAATCAAGAAAAGAA 11864
 QY 856 GTTAAATGATATTAATCTTTGTTGATCGTCGATCGTCAAGAAATTCATTGGGTGCA 915
 DB 11883 TTAAATTAATTAATTTTCAAGTGTGATGATATTTTAAACTATCTAATTTTGA 11824

QY 916 ACAAGATTCTCG 928
 DB 11823 AATATATATACCG 11811

RESULT 15
 AC096684 84472 bp DNA linear HTG 22-SEP-2001
 LOCUS AC096684
 DEFINITION Takifugu rubripes clone 241N7, WORKING DRAFT SEQUENCE, 3 unordered pieces.
 AC096684 GI:15721942
 VERSION AC096684.1 GI:15721942
 KEYWORDS HTG; HTGS PHASE1; HTGS_DRAFT.
 SOURCE Takifugu rubripes.
 ORGANISM Takifugu rubripes.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
 Tetraodontidae; Takifugu.

REFERENCE
 AUTHORS Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S., Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P., Ho, S.-L., Idol, J.R., Karlins, E., Latic, P., Lee-Lin, S.-O., Legaspi, R., Maduro, O.L., Maduro, V.B., Masillo, C., Mastrian, S.D., McCluskey, J.C., McDowell, J., Pearson, R., Prasad, A., Shevchenko, Y., Stantirip, S., Thomas, J.W., Thomas, P.J., Touchman, J.W., Tsugeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L., Zhang, L.-H. and Green, E.D.
 NISC Comparative Sequencing Initiative
 Unpublished
 2 (bases 1 to 84472)
 Green, E.D.
 DIRECT SUBMISSION
 Submitted (22-SEP-2001) NIH Intramural Sequencing Center, 8717 Grovemont Circle, Gaithersburg, MD 20877, USA

COMMENT
 Center: NIH Intramural Sequencing Center
 Center code: NISC
 Web site: http://www.nisc.nih.gov
 Contact: nisc.mouse@nigrl.nih.gov
 Project Information
 Center project name: arc
 Center clone name: 241N07

----- Summary Statistics
 Sequencing vector: plasmid; n/a; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 83639 bases at least Q40
 Consensus quality: 83810 bases at least Q30
 Consensus quality: 83800 bases at least Q20
 Insert size: 89000; agarose-fp
 Insert size: 101000; pulse-field-gel
 Insert size: 84272; sum-of-contigs
 Quality coverage: 12.81x in Q20 bases; agarose-fp
 Quality coverage: 11.29x in Q20 bases; pulse-field-gel
 Quality coverage: 13.53x in Q20 bases; sum-of-contigs

 * NOTE: This is a 'working draft' sequence. It currently consists of 3 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 10947: contig of 10947 bp in length
 * 10948 11047: gap of unknown length
 * 11048 37566: contig of 26519 bp in length
 * 37567 37666: gap of unknown length
 * 37667 84472: contig of 46806 bp in length.
 Location/Qualifiers

FEATURES

PI Shimkete RA, Leach MD;
 XX WPI; 2002-106308/14.
 DR P-PSDB; ABPI0927.
 XX
 XX Novel human polypeptides and polynucleotides useful for diagnosing,
 PT preventing and treating cardiovascular disease, neurodegenerative,
 PT hyperproliferative disorders and autoimmune disorders
 XX
 PS Disclosure; SEQ ID 21835; 1037pp; English.
 CC The present invention describes substantially purified human proteins
 CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1
 CC in the specification). ABN15762 to ABN27252 encode the human ORFX
 CC proteins given in ABP00010 to ABP1500. ORFX proteins are useful for
 CC treating or preventing a pathology associated with an ORFX-associated
 CC disorder in humans, and in the manufacture of a medicament for treating a
 CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide
 CC sequences can be used in gene therapy. ORFX sequences can be used in the
 CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
 CC psoriasis, benign tumors, keloid, degenerative disorders, haemorrhage,
 CC osteoarthritis, neurodegenerative disorders, disorders related to organ
 CC transplantation, cardiovascular diseases, diabetes mellitus, systemic
 CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester
 CC storage disease, various immune deficiencies and disorders, infectious
 CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
 CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
 CC disease and autoimmune inflammatory eye disease. ORFX proteins are also
 CC useful for treating burns, incisions, ulcers, for treating osteoporosis,
 CC bone degenerative disorders, or periodontal disease, and for gut
 CC protection or regeneration and treatment of lung or liver fibrosis,
 CC reperfusion injury in various tissues and conditions resulting from
 CC systemic cytokine damage.
 CC N.B. The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 XX
 SQ Sequence 393 BP; 101 A; 93 C; 91 G; 107 T; 1 other;
 Query Match 8.8%; Score 98.4; DB 24; Length 393;
 Best Local Similarity 55.8%; Pred. No. 9,7e-17;
 Matches 211; Conservative 0; Mismatches 161; Indels 6; Gaps 1;
 QY 225 TCGTAGTCTGACTGAGTGGAGTACCGGTTTGGTTTGTTCACGGTCTTCCCA 284
 DB 12 TGATGCTGCACGGCTTGCCATTGGTGAATACATTTAAGGGGTTATTCACGGTCTTCAGT 71
 285 AACAGATCTCAAAATGTCATTTGCTGAATATGCCAGGTTTATCCACCTTTGTTT 344
 72 GAGGTTCCCTCAACGCGTCTTGTGAGTACTGTCATTCATCCAGACTGCTT 131
 345 CAA-----GAGTAACTTAACCTCACTCACTGCTGATGAATTTCTGAAAGCCCTGTGA 398
 DB 132 TAAACCAAGCAGAGAGTTTATGATGTCGGTAAAGAACACTACCAAGAGCCCGTAA 191
 QY 399 GAACTTGCATCTGCTGATCATTTGCCAGTTTGTGACAACTGCTGTGTTGTTG 458
 DB 192 ATCTTTAGAAAGGCGAGTATCCCTCCAGCTCATTTGACACGCGTGTATGATCCCTTAC 251
 QY 459 TCATCACTTGGGCTCAAAATGGAATGAGCAGCCATGACCCGCAACATCTCATTCAT 518
 DB 252 ACATAGTTTGGCTTGGACATGACATGAAAGCCCTCAAGAGCGAAGAGATCAACCCAT 311
 QY 519 ATTGATTTGGGCTGTGCTCAAGCAGTGGGTCAACAACATAATCCAGTTGCCAAGCATAT 578
 DB 312 CTTATTTTGGGGTGTGCTGCTGTGTTGGCAGATGCTTATTCATTTGCAAAAAAACH 371
 QY 579 CAATGCTTATTAAT 596
 DB 372 AAACGGTTTCAGCAAGAT 389

AAV74349/C
 ID AAV74349 standard; DNA; 23439 BP.
 XX
 XX AAV74349;
 AC
 XX
 DT 16-MAR-1999 (first entry)
 XX
 XX Staphylococcus aureus contig SEQ ID #38.
 DE
 XX Computer readable medium; vaccine; S. aureus infection; immunodetection;
 KM cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
 KM skin infection; surgical wound infection; scaled skin syndrome;
 KM toxic shock syndrome; ds.
 XX
 OS Staphylococcus aureus.
 XX
 FH Key Location/Qualifiers
 FT misc_feature 481..540
 FT /tag= a
 FT /note= "these bases represent a line of missing text in
 FT the sequence listing in the specification. They
 FT are included to maintain the nucleotide numbering
 FT given in the specification for this DNA sequence"
 FT
 FT misc_feature 2281..2340
 FT /tag= b
 FT /note= "these bases represent a line of missing text in
 FT the sequence listing in the specification. They
 FT are included to maintain the nucleotide numbering
 FT given in the specification for this DNA sequence"
 FT
 FT misc_feature 4081..4140
 FT /tag= c
 FT /note= "these bases represent a line of missing text in
 FT the sequence listing in the specification. They
 FT are included to maintain the nucleotide numbering
 FT given in the specification for this DNA sequence"
 FT
 FT misc_feature 5881..5940
 FT /tag= d
 FT /note= "these bases represent a line of missing text in
 FT the sequence listing in the specification. They
 FT are included to maintain the nucleotide numbering
 FT given in the specification for this DNA sequence"
 FT
 FT misc_feature 7681..7740
 FT /tag= e
 FT /note= "these bases represent a line of missing text in
 FT the sequence listing in the specification. They
 FT are included to maintain the nucleotide numbering
 FT given in the specification for this DNA sequence"
 FT
 FT misc_feature 9481..9540
 FT /tag= f
 FT /note= "these bases represent a line of missing text in
 FT the sequence listing in the specification. They
 FT are included to maintain the nucleotide numbering
 FT given in the specification for this DNA sequence"
 FT
 FT misc_feature 11281..11340
 FT /tag= g
 FT /note= "these bases represent a line of missing text in
 FT the sequence listing in the specification. They
 FT are included to maintain the nucleotide numbering
 FT given in the specification for this DNA sequence"
 FT
 FT misc_feature 13081..13140
 FT /tag= h
 FT /note= "these bases represent a line of missing text in
 FT the sequence listing in the specification. They
 FT are included to maintain the nucleotide numbering
 FT given in the specification for this DNA sequence"
 FT
 FT misc_feature 14881..14940
 FT /tag= i
 FT /note= "these bases represent a line of missing text in
 FT the sequence listing in the specification. They
 FT are included to maintain the nucleotide numbering
 FT given in the specification for this DNA sequence"
 FT
 FT misc_feature 16681..16740
 FT /tag= j
 FT /note= "these bases represent a line of missing text in
 FT the sequence listing in the specification. They
 FT are included to maintain the nucleotide numbering
 FT given in the specification for this DNA sequence"

FT /note="these bases represent a line of missing text in
 the sequence listing in the specification. They
 are included to maintain the nucleotide numbering
 given in the specification for this DNA sequence"
 FT misc_feature 18481..18540
 FT /ctag=K
 FT /note="these bases represent a line of missing text in
 the sequence listing in the specification. They
 are included to maintain the nucleotide numbering
 given in the specification for this DNA sequence"
 FT misc_feature 20281..20340
 FT /ctag=1
 FT /note="these bases represent a line of missing text in
 the sequence listing in the specification. They
 are included to maintain the nucleotide numbering
 given in the specification for this DNA sequence"
 FT misc_feature 22081..22140
 FT /ctag=m
 FT /note="these bases represent a line of missing text in
 the sequence listing in the specification. They
 are included to maintain the nucleotide numbering
 given in the specification for this DNA sequence"
 XX EP786519-A2.
 XX 30-JUL-1997.
 XX 07-JAN-1997; 97BP-0100117.
 XX PF 05-JAN-1996; 96US-0009861.
 XX PR (HUMA-) HUMAN GENOME SCI INC.
 XX PA Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA;
 XX PI Rosen CA;
 XX WPI; 1997-374922/35.
 XX DR Polynucleotide(s) and proteins derived from *Staphylococcus aureus* -
 XX PT stored on computer readable medium and used in the production of
 XX PT anti-S. aureus vaccines
 XX PS Claim 1; Page 339-352; 3271bp; English.
 XX CC This sequence represents one of 5191 *Staphylococcus aureus* DNA sequences
 CC of the invention. The DNA sequences are recorded on a computer readable
 CC medium, preferably selected from a floppy or hard disk, random access
 CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
 CC the S. aureus DNA sequences allows putative functions to be assigned so
 CC that protein-encoding or regulatory regions of commercial, therapeutic or
 CC industrial importance can be obtained. Specifically, sequences which are
 CC likely to encode antigens have been identified and these polypeptides can
 CC be used in a vaccine composition against S. aureus infection. The
 CC polypeptides can also be used in a kit for the immunodetection of
 CC S. aureus in a sample. S. aureus is implicated in numerous human diseases,
 CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,
 CC skin and surgical wound infections, scaled skin syndrome, toxic shock
 CC syndrome, etc. Organisms transformed with the DNA sequences can be used
 CC for recombinant production of the polypeptides. The new DNA sequences
 CC (and their fragments) are useful as primers or probes for isolating
 CC homologues of any of the S. aureus DNA sequences contained on the
 CC computer readable medium.
 XX SQ Sequence 23439 BP; 7768 A; 3506 C; 4455 G; 6921 T; 789 other;
 OY Query Match 4.1%; Score 45.4; DB 18; Length 23439;
 Db Best Local Similarity 45.4%; Pred. No. 0.13;
 Matches 163; Conservative 0; Mismatches 196; Indels 0; Gaps 0;
 OY 504 ACATACCTCATTCATTTGGGCTGCTACAGAGTGGGTCAACACTATATCA 563
 Db 16078 ACACGCTTTTAACTACGTTATGACGCTGACAGGTAAATGATTAATATATTA 16019

OY 564 AGTTCACCAACATATCAATGCTTATCTAAGATGTAAGTGTCTTAAAGCATGA 623
 Db 16018 CATTCGATATATATATTAATTTAAATATATTTGACCAAGATGACTTAAACCTCC 15959
 OY 624 AAAGCTTTAAAGCTTATGCTGCTGATGATGCTTTGACTATCATGATGAGCGCTTAT 683
 Db 15958 ACCTGTTTCAATGCGCATTTGTAATGCGATTATGATGTCCTCCGTTGAGAAAGGACCA 15899
 OY 684 TGAGAGATCAATGGAAGTATCCAAACCTGCAATGTTATTTAGCGCTGGGAAGCA 743
 Db 15898 TATCATGATCAATGTTCCACATTCGAAACGCTGATATAAACGAGCATGATTT 15839
 OY 744 AGATAGTATCCCGAGGCTTAAAGTCACAGCATAGTCTTACCTGCGCATTTATGA 803
 Db 15838 TGATGTGAGCTTATATGACGAAAGCATATCATTTATGTTAAAGTTGATTTCTTAA 15779
 OY 804 AGTGTTCATGACCATTTGAAGCATTCCTGAAAGAAATCAGAAAGATATGTTAA 862
 Db 15778 CGCTATTAACAATGATGATATATCATATTTTAAACCAATTATTCATCATGATA 15720
 RESULT 3
 AA70202
 ID AAA70202 standard; DNA; 3501 BP.
 AC AAA70202;
 XX 07-NOV-2000 (first entry)
 DT 07-NOV-2000 (first entry)
 XX DE Plasmodium falciparum chromosome 2 related DNA sequence SEQ ID NO:335.
 XX KW Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
 XX KW antimalarial; malaria; protozoacide; infection; insecticide; ds.
 XX OS Plasmodium falciparum.
 XX PN WO200025728-A2.
 XX PD 11-MAY-2000.
 XX PF 05-NOV-1999; 99WO-US26796.
 XX PR 05-NOV-1998; 98US-0107131.
 XX PA (HOFF/) HOFFMAN S.
 XX PA (CARU/) CARUCCI D.
 XX PA (GARD/) GARDNER M.
 XX PA (VENT/) VENTER J C.
 XX PI Hoffman S, Carucci D, Gardner M, Venter JC;
 XX DR WPI; 2000-365347/31.
 XX PT Proteins encoded by chromosome 2 of the human malarial parasite,
 XX PT Plasmodium falciparum, useful as antimalarial vaccines and in the
 XX PT diagnosis of P. falciparum infection -
 XX PS Disclosure; Page 528-529; 577bp; English.
 XX CC The present invention describes proteins and their fragments (I) encoded
 CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
 CC Also described are: (1) nucleotide sequences (II) encoding (I); and (2)
 CC vaccines against P. falciparum infection comprising (I) or (II).
 CC (I) and (II) are useful for the development of vaccines against
 CC P. falciparum infection. (I) and polyclonal antisera or a monoclonal
 CC antibody raised to immunogens comprising the sequences of (I), are
 CC useful in the detection of infection with P. falciparum. Furthermore,
 CC (I) (especially when they are rifins or secreted or membrane proteins)
 CC can aid the identification of drugs to treat or prevent P. falciparum
 CC infection, or they can be used to identify drug resistance in
 CC P. falciparum. Sequencing of the Plasmodium chromosome 2 and the
 CC subsequent identification of proteins encoded by it will help to expand
 CC our understanding of parasite biology, a process hampered by the

CC complex of the parasitic lifecycle, and provide new targets for
CC vaccine and drug development. Parasite resistance to drugs and mosquito
CC resistance to insecticides have led to a resurgence of malaria in many
CC parts of the world, and there is a pressing need for vaccines and new
CC drugs. AA07078 to AA070287 and AB01844 to AB018352 represent nucleotide
CC and protein sequences given in the present invention, but which are not
CC specifically mentioned within the specification.

SQ Sequence 3501 BP; 1699 A; 327 C; 468 G; 1007 T; 0 other;

Query Match	3.7%	Score 40.8	DB 21	Length 3501
Best Local Similarity	46.2%	Pred. No. 1.1		
Matches 209, Conservative	0	Mismatches 237	Indels 6	Gaps 2

Oy		650	TGATGTCCTTTGCATCATCATGATGACGGCGTATTAGACAATCAAAATCGAAGTATCCAA	709
Dd		194	ATGATGAATGTGTTTAAGACGACGTCGTAAGTGATGAGAGATCCACMAAGAAAAACGA	253
Oy		710	ACCTGCACAATGTTATTGACCGCTGTGGAGACGAAGATGATGCCCGAGCGCATTAAGA	769
Dd		254	ACCATGATTAATATAAACATATATGATGTAATGATGATTAATAT--AAATATATATCCTA	310
Oy		770	TCAACGCGATATGTTACTCTGCCACATTTATAGAAAGTGTTCCATGACCATTTGAAGCA	829
Dd		311	TAAATGATGATTAATTAATGAACAATATCATATATGATGATGATTAATTAACAATATACATA	370
Oy		830	TTCCGCGAAGAAATCGAAAAGATATGTTTAAATATGATATTACTTTGTTGATCGTGCAAT	889
Dd		371	TAAATGATGATTAATTAACATATATGATGTAAGATGATTAATTAATTAATATATCATTT	430
Oy		890	CTGCTCAAGAAATTCATTTGGGTGCACACAGATTTCTCTGTCGTCCAGATATCATGAAAG	949
Dd		431	TAAATGATGATTAATTAATTAATATATCATATTAATGAAGATTAATTAATTAATTAACCTATA	490
Oy		950	CCACGTTAAATTCGTTAAGTTTAAATTAATCCAACCTTAACAAGSGATATCCATCATATA	1009
Dd		491	TBAATGATGATTAATTAACATTAACATTAATTAAGATGATTAATTAATTAATTAATTTATT	550
Oy		1010	TGAATATTAAAGTTTTTCAGCAACGGCTTAGATG---ATGTCCACAGCTCTCACAGAGTA	1066
Dd		551	GTAATTAATGATGATTAATGATTAACATATCAATCAAAATGTTCAATATCATCTGTAATATTTCTCTTA	610
Oy		1067	TAAAGAAAGTTRAAAAACAAAATGTTAAAGTAT	1098
Dd		611	AAAAAAAAAAAAAGAAAAATGTAATAT	642

JULY 4

ABA49878 standard; DNA; 439 BP.

AC ABA49878;

DT 01-FEB-2002 (first entry)

DE Human breast cell single exon nucleic acid probe #8573:

KW Human; microarray; single exon probe; gene expression; breast;

• • • • •

OS Homo sapiens.

PN WO200157271-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US006662.

PR 04-FEB-2000; 2000US-0180312.

PR 30-JUN-2000; 2000US-0608408.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.

PA (MOLE-) MOLECULAR DYNAMICS 1

PI penn SG, Hanzel DK, Chen W,

DR WPI; 2001-496933/54.

PT New spatially-addressed

PT New spatially-addressable set of single exon nucleic acid probes,
PT useful for measuring gene expression in sample derived from human
PT breast, comprises number of single exon nucleic acid probes -
XX
PS Claim 4, SEQ ID NO 8573; 327pp + sequence listing; English.

Claim 4; SEQ ID NO 8573; 327pp + sequence listing; English

The invention relates to a spatially-addressable set of single exon nucleic acid probes for measuring gene expression in a sample derived from human breast and BT 474 cells. The method involves contacting the probes with a collection of detectably labelled nucleic acids derived from mRNA of human breast, and then measuring the label bound to each probe of the microarray. The probes are useful for verifying the expression of regions of genomic DNA predicted to encode proteins. They are useful for gene discovery, and for determining predisposition and/or prognosing breast disease. Gene expression analysis is useful for assessing the toxicity of chemical agents on cells. The microarray of this invention presents a far greater diversity of probes for measuring gene expression, with far less bias than expressed sequence tag microarrays. The method is suitable for rapid production of functional information from genomic sequence. The present sequence is a single exon nucleic acid probe of the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO. http://ftp.wipo.int/pub/published_pct_sequences.

Sequence 439 BP; 141 A; 169 C; 41 G; 88 T; 0 other;

Query Match	3.6%	Score 40.2	DB 22	Length 439
Best Local Similarity	46.8%	Pred. No. 0.66		
Matches 126	Conservative 0	Mismatches 143	Indels 0	Gaps 0

[illegible]

RESULT 5

ID ABA67796 standard; DNA; 439 BP.

AC ABA67796;

DT 01-FEB-2002 (first entry)

Human foetal liver single exon nucleic acid probe #16101.

Human; foetal liver; gene expression; single exon nucleic acid probe; ss.

OS Homo sapiens.
 XX KW MO200157277-A2.
 XX PD 09-AUG-2001.
 XX PF 30-JAN-2001; 2001WO-US00669.
 XX PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX PF WPI, 2001-483447/52.
 XX PT Human genome-derived single exon nucleic acid probes useful for
 XX PT analyzing gene expression in human fetal liver -
 XX PS Claim 4; SEQ ID NO 16101; 639pp + sequence listing; English.
 XX CC The invention relates to a single exon nucleic acid probe for
 CC measuring human gene expression in a sample derived from human foetal
 CC liver. The single exon nucleic acid probes may be used for predicting,
 CC measuring and displaying gene expression in samples derived from human
 CC fetal liver. The present sequence is a single exon nucleic acid
 CC probe of the invention.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 XX SQ Sequence 439 BP; 141 A; 169 C; 41 G; 88 T; 0 other;
 XX
 Query Match 3.6%; Score 40.2; DB 22; Length 439;
 Best Local Similarity 46.8%; Pred. No. 0.66;
 Matches 126; Conservative 0; Mismatches 143; Indels 0; Gaps 0;
 QY 36 TGAAGTGCAAGCTGTTGTTAAACAGATGTCAGTCCAGATTAAAGAGGGTAC 95
 DB 385 TGATGTGTGTGATGTGTGATGCTGCTGCTGATGATGATGCGGATTAATGATGATGTC 326
 QY 96 AGCTTGTGTAAGTTGAGGCTGTTGCTGTAACCACTGATGGAAGCATATTGCTTA 155
 DB 325 TCCGATGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 266
 QY 156 TAAAGTTGTCAGAGGTTCAATTCTAGATGTACATGCTGCTGATGATGATGATGATG 215
 DB 265 TGCATGATGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 206
 QY 216 TGAACCAATGCTAGTACTGATGATGATGATGATGATGATGATGATGATGATGATG 275
 DB 205 TCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 146
 QY 276 TGCCTCCCAACAGATCTCTAAATAATGCTG 304
 DB 145 TGATGATGATGATGATGATGATGATGATG 117
 RESULT 6
 ID ABA34854 standard; DNA; 439 BP.
 AC ABA34854;
 XX 23-JAN-2002 (first entry)
 DT
 XX Probe #13320 for gene expression analysis in human heart cell sample.
 DE

XX KW Human; gene expression; heart; microarray; vascular system; probe;
 KW cardiovascular disease; hypertension; cardiac arrhythmia;
 KW congenital heart disease; ss.
 XX OS Homo sapiens.
 XX KW MO200157274-A2.
 XX PD 09-AUG-2001.
 XX PF 30-JAN-2001; 2001WO-US00666.
 XX PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX PF WPI, 2001-488899/53.
 XX DR WPI, 2001-488899/53.
 XX CC Single exon nucleic acid probes for analyzing gene expression in human
 XX CC hearts -
 XX PS Claim 4; SEQ ID NO 13320; 530pp; English.
 XX CC The present invention relates to single exon nucleic acid probes for
 CC measuring human gene expression in a sample derived from human heart. The
 CC present sequence is one such probe. The probes may be used for
 CC predicting, measuring and displaying gene expression in samples derived
 CC from the human heart via microarrays. By measuring gene expression, the
 CC probes are useful for predicting, diagnosing, grading, staging,
 CC monitoring and prognosing diseases of the human heart and vascular system
 CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
 CC congenital heart disease.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX SQ Sequence 439 BP; 141 A; 169 C; 41 G; 88 T; 0 other;
 XX
 Query Match 3.6%; Score 40.2; DB 22; Length 439;
 Best Local Similarity 46.8%; Pred. No. 0.66;
 Matches 126; Conservative 0; Mismatches 143; Indels 0; Gaps 0;
 QY 36 TGAAGTGCAAGCTGTTGTTAAACAGATGTCAGTCCAGATTAAAGAGGGTAC 95
 DB 385 TGATGTGTGTGATGTGATGCTGCTGCTGATGATGATGATGATGATGATGATGATG 326
 QY 96 AGCTTGTGTAAGTTGAGGCTGTTGCTGTAACCACTGATGGAAGCATATTGCTTA 155
 DB 325 TCCGATGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 266
 QY 156 TAAAGTTGTCAGAGGTTCAATTCTAGATGTACATGCTGCTGATGATGATGATGATG 215
 DB 265 TGCATGATGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 206
 QY 216 TGAACCAATGCTAGTACTGATGATGATGATGATGATGATGATGATGATGATGATG 275
 DB 205 TCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 146
 QY 276 TGCCTCCCAACAGATCTCTAAATAATGCTG 304
 DB 145 TGATGATGATGATGATGATGATGATGATG 117
 RESULT 7

Db 145 TGATGATGATGTAATGATGATGATG 117

RESULT 9

AA122718/c
ID AA122718 standard; DNA; 439 BP.

AC AA122718;

DT 12-OCT-2001 (first entry)

DE Probe #12651 for gene expression analysis in human cervical cell sample.

KW Probe; human; microarray; gene expression; cervical epithelial cell;

OS Homo sapiens.

PN WO200157278-A2.

09-AUG-2001.

30-JAN-2001; 2001WO-US00670.

04-FEB-2000; 2000US-0180312.

26-MAY-2000; 2000US-0207456.

30-JUN-2000; 2000US-0608408.

03-AUG-2000; 2000US-0632366.

21-SEP-2000; 2000US-0234687.

27-SEP-2000; 2000US-0236359.

04-OCT-2000; 2000GB-0024263.

(MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

WPI; 2001-488901/53.

Human genome-derived single exon nucleic acid probes useful for

analyzing gene expression in human cervical epithelial cells -

Claim 25; SEQ ID No 12651; 487bp; English.

The present invention relates to human single exon nucleic acid probes (SENP). The present sequence is one such probe. The SENPs are derived from human HeLa cells. The SENPs can be used to produce a single exon microarray, which can be used for measuring human gene expression in a sample derived from human cervical epithelial cells. By measuring gene expression, the probes are therefore useful in grading and/or staging of diseases of the cervix, notably cervical cancer. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp://wipo.int/pub/published_pct_sequences](http://wipo.int/pub/published_pct_sequences).

Sequence 439 BP; 141 A; 169 C; 41 G; 88 T; 0 other;

Query Match 3.6%; Score 40.2; DB 22; Length 439;

Best Local Similarity 46.8%; Pred. No. 0.66; Mismatches 143; Indels 0; Gaps 0;

Matches 126; Conservative 0; Mismatches 143; Indels 0; Gaps 0;

36 TGAAGTGCACAAAGCTGTTTAAACAGATGTCAGTTCAGATTAAAGAGAGGTAC 95

385 TGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 326

96 AGCTTGTGTAAGGTTGAGGCTGTTGCTGTAACCAACTGATGGAAGCATATTGCTTA 155

325 TGCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 266

156 TAAAGTTGTCAGAAAGGTTCAATTCTAGATGATGATGATGATGATGATGATGATG 215

265 TCGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 206

216 TGAACCAATGCTAGTACTGATGTAAGGTTGAGATACCGGTTTCGTTTGTTCACGG 275

Db 205 TCCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 146

QY 276 TGCCTCCCAACAGATCCTTAATAATGCTG 304

Db 145 TGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 117

RESULT 10

AA148016/c
ID AA148016 standard; DNA; 439 BP.

AC AA148016;

DT 17-OCT-2001 (first entry)

DE Probe #16702 used to measure gene expression in human placenta sample.

KW Probe; microarray; human; placenta; antenatal diagnosis;

OS Homo sapiens.

PN WO200157272-A2.

09-AUG-2001.

30-JAN-2001; 2001WO-US00663.

04-FEB-2000; 2000US-0180312.

26-MAY-2000; 2000US-0207456.

30-JUN-2000; 2000US-0608408.

03-AUG-2000; 2000US-0632366.

21-SEP-2000; 2000US-0234687.

27-SEP-2000; 2000US-0236359.

04-OCT-2000; 2000GB-0024263.

(MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

WPI; 2001-488897/53.

Human genome-derived single exon nucleic acid probes useful for

analyzing gene expression in human placenta -

Claim 25; SEQ ID No 16702; 654bp; English.

The present invention relates to single exon nucleic acid probes (SENP). The present sequence is one such probe. The probes are useful for producing a microarray for predicting, measuring and displaying gene expression in samples derived from human placenta. The probes are useful for antenatal diagnosis of human genetic disorders.

Sequence 439 BP; 141 A; 169 C; 41 G; 88 T; 0 other;

Query Match 3.6%; Score 40.2; DB 22; Length 439;

Best Local Similarity 46.8%; Pred. No. 0.66; Mismatches 143; Indels 0; Gaps 0;

Matches 126; Conservative 0; Mismatches 143; Indels 0; Gaps 0;

36 TGAAGTGCACAAAGCTGTTTAAACAGATGTCAGTTCAGATTAAAGAGAGGTAC 95

385 TGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 326

96 AGCTTGTGTAAGGTTGAGGCTGTTGCTGTAACCAACTGATGGAAGCATATTGCTTA 155

325 TGCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 266

156 TAAAGTTGTCAGAAAGGTTCAATTCTAGATGATGATGATGATGATGATGATGATG 215

265 TCGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 206

216 TGAACCAATGCTAGTACTGATGTAAGGTTGAGATACCGGTTTCGTTTGTTCACGG 275

CC mRNA, and (b) measuring the label detectably bound to each probe of
 CC the array; identifying exons in a eukaryotic genome, comprising
 CC (a) algorithmically predicting at least one exon from genomic sequences
 CC of the eukaryote; and (b) detecting specific hybridisation of detectably
 CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,
 CC having a fragment identical to the predicted exon, the probe is included
 CC in the above mentioned microarray; assigning exons to a single gene,
 CC comprising (a) identifying exons from genomic sequence by the method
 CC above and (b) measuring the expression of each of the exons in several
 CC tissues and/or cell types using hybridisation to a single exon
 CC microarrays having a probe with the exon, where a common pattern of
 CC expression of the exons in the tissues and/or cell types indicates that
 CC the exons should be assigned to a single gene; a peptide comprising one
 CC of 12011 sequences, mentioned in the specification, or encoded by the
 CC probes/open reading frames (ORF). The probes are used for gene
 CC expression analysis, and for identifying exons in a gene, particularly
 CC using human lung derived mRNA and for the study of lung diseases
 CC such as asthma, lung cancer, chronic obstructive pulmonary disease
 CC (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary
 CC fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,
 CC Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary
 CC haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomatosis,
 CC pulmonary alveolar proteinosis, Kartagener syndrome, fibrocystic
 CC pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension
 CC and hyaline membrane disease. The present sequence is a single exon
 CC probe. Open reading frame of the invention.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.

CC Sequence 439 BP; 141 A; 169 C; 41 G; 88 T; 0 other;

CC Query Match 3.6%; Score 40.2; DB 24; Length 439;
 CC Best Local Similarity 46.8%; Pred. No. 0.66;
 CC Matches 126; Conservative 0; Mismatches 143; Indels 0; Gaps 0;

CC 36 TGAAGCTGCAAGAGCTGTTTAAACAGATGTCAGTCCAGATTAAAGAGGGTAC 95
 CC 35 TGATGTGTGTATGTGTATGCTGCTGCTGATATATGTTGGGATATGATGATGATC 326
 CC 96 AGCTTGTGTAAGGCTGTTGCTGCTGTTGTAACCACTGATGTAAGCATATTTGCTTA 155
 CC 325 TGCCATGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 266
 CC 156 TAAAGTTGTGCAAGAGCTTCAATCTAGAGATGATGATGATGATGATGATGATG 215
 CC 265 TGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 206
 CC 216 TGACCAATGCTAGTACTGATGATGATGATGATGATGATGATGATGATGATGATG 275
 CC 205 TCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 146
 CC 276 TGCTTCCCAACAGATCTTAAATGATG 304
 CC 145 TGATGATGATGATGATGATGATGATG 117

CC RESULT 13

CC ABL56207 standard; DNA; 1050 BP.

CC ABL56207;

CC 01-JUL-2002 (first entry)

CC AMBPV first RNA polymerase (AMV051) encoding sequence.

CC AMBPV; gene therapy; viral vector; chromosome mapping; gene mapping;
 CC genetic deficiency disorder; gene; RNA polymerase; ds.

CC Amsacta moorei entomopoxvirus.

CC Key Location/Qualifiers
 CC FT CDS 1..1050
 CC FT /tag= a
 CC FT /product= "first RNA polymerase (AMV051)"

CC MO200212526-A2.

CC 14-FEB-2002.

CC 10-AUG-2001; 2001WO-US25287.

CC 10-AUG-2000; 2000US-224479P.

CC 14-SEP-2000; 2000US-0662254.

CC (UNIV FLORIDA.

CC Moyer RW, Li Y, Bawden AL;

CC WPI; 2002-227161/28.

CC P-PSDB; ABB77604.

CC Novel recombinant entomopox virus vector useful for delivering
 CC polynucleotide encoding protein to vertebrate cell, comprises
 CC polynucleotide encoding protein operably linked with heterologous
 CC promoter sequence -

CC Claim 74; Page 245-247; 326pp; English.

CC The invention relates to a recombinant entomopox virus (BPV) vector,
 CC comprising a polynucleotide encoding a protein operably linked with a
 CC heterologous promoter sequence. The invention also concerns methods for
 CC providing gene therapy for genetic deficiency disorders. Vectors of the
 CC invention are useful for delivering a polynucleotide encoding a protein
 CC to a vertebrate cell preferably a mammalian cell, such as a human cell.
 CC The vector is introduced into the vertebrate cell by infection in a viral
 CC particle, or by transfection, transduction, or injection either in vitro
 CC or in vivo. The vector is useful for the delivery and expression of
 CC biologically useful proteins in gene therapy protocols, and for
 CC delivering large DNA segments for engineering of vertebrate cells.
 CC Polynucleotides of the invention have applications in techniques such as
 CC their use as insertion sites for foreign genes of interest, hybridisation
 CC probes, for chromosome and gene mapping, in PCR technologies, and in the
 CC production of sense or antisense nucleic acids. Vectors of the invention
 CC provide for stable integration and expression of heterologous DNA in host
 CC cells, and are adapted for accepting large heterologous polynucleotide
 CC inserts which can be delivered in an infected or transformed cell and
 CC expressed in a stable fraction. The current sequence represents an
 CC amsacta moorei entomopoxvirus (AMBPV) first RNA polymerase (AMV051)
 CC encoding sequence.

CC Sequence 1050 BP; 487 A; 85 C; 111 G; 367 T; 0 other;

CC Query Match 3.6%; Score 39.6; DB 24; Length 1050;
 CC Best Local Similarity 49.5%; Pred. No. 1.4;
 CC Matches 102; Conservative 0; Mismatches 104; Indels 0; Gaps 0;

CC 889 TCTGCTGAGAAATCTATTTGGTGCAACAGATTTCTGCTAGTCCAGATATCATG 948
 CC 559 TATTATGAAAAAATTTATATATATAAACCATTCTTATGATAGCAAAATATCATATA 618
 CC 949 GCCACAGTTAAATTCGTTAAGTTTAAATCCACACTTAAACAGCGATATCCATCAT 1008
 CC 619 GAAAAAGATATTTATACATATGATATATATATATATATATATATATATATATAT 678
 CC 1009 ATGAATATTAAGTTTTCAGCAACGCGCTTAAGATGTCACAGCTTCATGAGATATA 1068
 CC 679 AATATATAAAAAATCTTATGATATATATATATAAATAGACATATATTTGAAAAATAT 738
 CC 1069 AAAGAGGTAATAAACAATAATGTTAA 1094
 CC 739 AAAATTAATAAATAATATGTTGATTA 764

RESULT 14
AAK67473
ID AAK67473 standard; DNA; 4027 BP.
XX
AC AAK67473;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:22285.
XX
DE Human immune/haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; dr.
XX
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
17-JAN-2001; 2001WO-US01354.
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180658.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217487.
PR 14-JUL-2000; 2000US-0218280.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.

PR 14-SEP-2000; 2000US-0233397.
PR 14-SEP-2000; 2000US-0233398.
PR 14-SEP-2000; 2000US-0233399.
PR 14-SEP-2000; 2000US-0233400.
PR 14-SEP-2000; 2000US-0233401.
PR 14-SEP-2000; 2000US-0233403.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.

PA	(HUMA-) HUMAN GENOME SCI INC.
XX	Rosen CA, Barash SC, Ruben SW,
XX	WPI; 2001-483426/52.
PT	Nucleic acids encoding human immune/hematopoietic antigen polypeptides, useful for preventing, diagnosing and/or treating cancers and metastasis -
PS	Disclosure; SEQ ID NO 22285; 3071bp + Sequence Listing; English.
CC	AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
CC	amino acid sequences given in AAM82170 to AAM91921. (I) have cytosolic
CC	activity, and can be used in gene therapy and vaccine production. (I)
CC	proteins and polynucleotides may be used in the prevention, diagnosis and
CC	treatment of diseases associated with inappropriate (I) expression. For
CC	example, they may be used to treat disorders associated with decreased
CC	expression by rectifying mutations or deletions in a patient's genome
CC	that affect the activity of (I) by expressing inactive proteins or to
CC	supplement the patients own production of (I). Additionally, (I)
CC	polynucleotides may be used to produce the secreted (I), by inserting
CC	the nucleic acids into a host cell and culturing the cell to express the
CC	protein. (I) proteins and polynucleotides may be used to prevent,
CC	diagnose and treat immune/hematopoietic-related diseases, especially
CC	cancers and cancer metastases of hematopoietic-derived cells. AAK64703
CC	to AAK87694 represent human immune/hematopoietic antigen genomic
CC	sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC	represent sequences used in the exemplification of the present invention.
XX	
SQ	Sequence 4027 BP, 1247 A, 875 C; 797 G; 1108 T; 0 other;
Query Match	3.6%; Score 39.6; DB 22; Length 4027;
Best Local Similarity	44.0%; Pred. No. 2.4;
Matches 168; Conservative	0; Mismatches 214; Indels 0; Gaps 0
Db	639 TTATGTCCTGATGATGTCTTTGACTATCATATGACAGCGCTTATTGACAGATCAATC 698
	987 TTCGACGATTTCTTGTCCTCAGCCTCCCAAGTAGCTGGAAATTACAGGTGTGCACCA 1044
Oy	699 GAAGTATCCAACCTGCMAAATGTATTAAGCGCTGTGGGAACGAAGATGATGCCGA 758
Ddb	1047 CACTCAGCTAATTTTTGTATTTTGTATTTGATGAGGCAAGGGTTTSCGAGTTAACCAAGGCTGG 1100
Oy	759 GGCCATAAAGTCACAGCAGATAGTCTACCTGCCACATTTATAGAAGTGGTTCATGAC 818
Ddb	1107 TCTCGAATCTCGACCTCAGAGTGATCCGCTGCTCGCTCCCAAAGTCGGGATTAC 1166
Oy	819 CATGGAANCAATTCCTGAAGAAATCAGAAAAGATTAATGTTAAATGATTAATCTTGT 878
Ddb	1167 AGCGCTGAGCACCAACAACCAGCCTATTAACAATTTGTTCTAATTAAGAAATGTGAAC 1224
Oy	879 GTATGTCGATCTGCTCAAGAAATTCATATGGGTGCAACAATTTCCGCTAGTCCACA 938
Ddb	1227 GAGAAAAACATTTAAGAAAGTATCAAAATTTAGTCAAAAGGACCTTCGAGATTCACT 1288
Oy	939 ATATCATGAAGCCACAGTTAAATTCGTTAAGTTTAAATTCACACCTTAAACAAGGTGA 998
Ddb	1287 TACTTAAAGTGAATTAATGATTAATTTATTTCTTCAAAACACGAAACCTGGGTTA 1344
Oy	999 TATCCATCATATGAATATTAAA 1020

[illegible]

```
QY 949 GCCACGTTAAATTCGTTAGTTAATCAACACCTTAACAAGGTGATATCCATCAT 1008
Db 1559 GAAAAAGATATTATTAATCTATAGATATAGATATATCATATATGATATTCCTAAAGAA 1500
QY 1009 ATGAATATTAAAGTTTTCAGCAACGGCTTAGATGATGCCAGCTCTCACTGAAGGATA 1068
Db 1499 AATATTAATAAATTCCTTATTTGATATATTTAATAAATAGCAGATATATTTGAAAAATAT 1440
QY 1069 AAAGAAAGTAAAAACAAAAATGTTAA 1094
Db 1439 AAAATTAAAAAATAATGTTGATTA 1414
```

Search completed: June 17, 2003, 07:59:16
Job time : 305 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 06:31:52 ; Search time 66 Seconds
(without alignments)
5171.685 Million cell updates/sec

Title: US-10-081-644-1
Perfect score: 1113
Sequence: 1 atgcagctcccaacacatca.....agtatgtccaggttataa 1113

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues
1 number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	37.6	3.4	8654	US-08-961-527-98	Sequence 98, Appl
2	35.8	3.2	840	US-09-042-771-1	Sequence 1, Appl
3	35.6	3.2	3600	US-08-855-910-7	Sequence 7, Appl
4	35.6	3.2	3792	US-08-992-334-1	Sequence 1, Appl
5	35.6	3.2	3792	US-08-302-752-1	Sequence 1, Appl
6	35.6	3.2	5234	US-08-992-334-2	Sequence 2, Appl
7	35.6	3.2	5234	US-08-302-752-2	Sequence 2, Appl
8	35.6	3.2	6722	US-08-992-334-3	Sequence 3, Appl
9	35.6	3.2	6722	US-08-302-752-3	Sequence 3, Appl
10	35.2	3.2	1434	US-09-134-001C-650	Sequence 650, App
11	34.8	3.1	1044	US-09-134-001C-461	Sequence 461, App
12	34.8	3.1	1347	US-08/622	INFORMATION FOR
13	34.8	3.1	1347	US-09-165-922A-11	Sequence 11, Appl
14	34.8	3.1	4016	US-09-173-053-3	Sequence 3, Appl
15	34.8	3.1	51259	US-08-781-891-209	Sequence 209, App
16	34.4	3.1	1856	US-08-360-606B-29	Sequence 29, App
17	34.2	3.1	1670	US-09-026-482B-1	Sequence 1, Appl
18	34.2	3.1	3828	PCT-US93-10580-1	Sequence 1, Appl
19	34.2	3.1	7721	US-08-772-270A-14	Sequence 14, Appl
20	34	3.1	15894	US-08-348-891A-1	Sequence 1, Appl
21	34	3.1	15894	US-08-905-817-1	Sequence 1, Appl
22	33.4	3.0	1074	US-09-134-001C-382	Sequence 382, App
23	33.4	3.0	1242	US-09-705-299-12	Sequence 12, Appl
24	33.4	3.0	12597	US-08-785-071A-1	Sequence 1, Appl
25	33.2	3.0	2631	US-09-012-872-1	Sequence 1, Appl
26	33.2	3.0	2631	US-08-930-996A-1	Sequence 1, Appl
27	33.2	3.0	4946	US-08-930-996A-1	Sequence 1, Appl

28	33.2	3.0	41708	US-09-470-512A-3	Sequence 3, Appl
29	33	3.0	384	US-09-325-932A-26	Sequence 26, Appl
30	33	3.0	5134	US-08-310-912A-157	Sequence 157, App
31	33	3.0	5134	US-09-301-085-157	Sequence 157, App
32	33	3.0	5134	PCT-US95-04589-157	Sequence 157, App
33	32.8	2.9	301	US-09-605-785-278	Sequence 278, App
34	32.8	2.9	301	US-09-439-313-278	Sequence 278, App
35	32.8	2.9	301	US-09-352-616A-278	Sequence 278, App
36	32.8	2.9	301	US-09-232-149A-278	Sequence 278, App
37	32.8	2.9	2031	US-09-693-147-5	Sequence 5, Appl
38	32.8	2.9	3095	5231168-1	Patent No. 5231168
39	32.8	2.9	7425	US-09-453-702B-212	Sequence 212, App
40	32.6	2.9	2234	US-09-221-017B-1062	Sequence 1062, App
41	32.6	2.9	2234	US-08-488-706-3	Sequence 3, Appl
42	32.6	2.9	4565	US-08-776-265-1	Sequence 1, Appl
43	32.6	2.9	5718	US-08-714-918-48	Sequence 48, Appl
44	32.6	2.9	5718	US-09-265-315-48	Sequence 48, Appl
45	32.6	2.9	5718	US-09-232-315-48	Sequence 48, Appl

ALIGNMENTS

RESULT 1
US-08-961-527-98

Sequence 98, Application US/08961527

Patent No. 6420135

GENERAL INFORMATION:

APPLICANT: Charles Kunsch

TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences

NUMBER OF SEQUENCES: 391

CORRESPONDENCE ADDRESS:

ADDRESS: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: Maryland

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

COMPUTER: HP Vectra 486/33

OPERATING SYSTEM: MSDOS version 6.2

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/961,527

FILING DATE:

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Brookes, A. Anders

REGISTRATION NUMBER: 36,373

REFERENCE/DOCKET NUMBER: PB340P1

TELECOMMUNICATION INFORMATION:

TELEPHONE: (301) 309-8504

TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 98:

SEQUENCE CHARACTERISTICS:

LENGTH: 8654 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

US-08-961-527-98

Query Match

Best Local Similarity 54.3%; Pred. No. 1.2;

Matches 76; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 788 CTGCACACTTATTGAAGTGTTCACATTTGAAGCATTCCTGAAGATTCAGAA 847
DB 6452 CTTCACATCTGTGTAGAGTGTGCGAGATTGATGATCTATTGAACTGGAATCCGTCG 6511

QY 848 AACATAATGTTAAATGATATTACTTTGTTATGTCGATCTGTCAGAAATTCAT 907
DB 6512 AACATGATATCAAGATGATACCTTCGCTTCAGGTGTGCGGTGACAAACGTCATA 6571
QY 908 TGGGTGCAACAGATTTCCT 927
DB 6572 AGGTTCAACAGGTGTCGT 6591

RESULT 2

US-09-042-771-1
; Sequence 1, Application US/09042771
; Patent No. 6080729

GENERAL INFORMATION:

APPLICANT: Jaworski, Deborah
APPLICANT: Wang, Min
APPLICANT: Shilling, Lisa K.
APPLICANT: Lawlor, Elizabeth
TITLE OF INVENTION: sp002
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:

ADDRESS: 4000 Bell Atlantic Tower, 1717 Arch St

CITY: Philadelphia
STATE: PA
COUNTRY: US
ZIP: 19103

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: FastSeq for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/042,771
FILING DATE:
CLASSIFICATION:

PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Falk, Stephen T
REGISTRATION NUMBER: 36,795
REFERENCE/DOCKET NUMBER: GM10135
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-994-2488
TELEFAX: 215-994-2222
TELEX:

INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 840 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-09-042-771-1

Query Match 3.2%; Score 35.8; DB 3; Length 840;
Best Local Similarity 51.6%; Pred. No. 1.5;
Matches 82; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

QY 281 CCCAACAGATCTCTAAATATGTCATTTGCTGATATGTCAGGCTTTATCCACCTTTG 340
DB 290 CTCACAGACGCTTATTTATTCGTATGATGATGATGATGATGATGATGATGATGAT 349
QY 341 TTACAGAGTAATCTTAATCACTCACTGCTGATGATGATGATGATGATGATGATGATGAT 400
DB 350 TTGAGATATTTAAAGGAAATTTGCTGTTGTTGAAGAGCGAAGCTATTAAGAAAT 409
QY 401 ACTTGCAATCTGTCATCATTCGCCAGTTTCGTTGCAA 439
DB 410 TATTGAAATTTGTTGATACCAACCAAGTGAATTTGCAA 448

RESULT 3

US-08-855-910-7
; Sequence 7, Application US/08855910
; Patent No. 6221640

GENERAL INFORMATION:

APPLICANT: Tao, Jianhui
APPLICANT: Sasanfar, Mandana
APPLICANT: Gallant, Paul L.
APPLICANT: Shen, Xiaoyu
APPLICANT: Avruch, Anthony S.
APPLICANT: Yu, Russell V.
APPLICANT: Nair, Shamila
TITLE OF INVENTION: ENTEROCOCCAL AMINOACYL-tRNA SYNTHETASE
NUMBER OF SEQUENCES: 71
CORRESPONDENCE ADDRESS:
ADDRESS: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Millita Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: USA
ZIP: 02173

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/855,910
FILING DATE: 14-MAY-1997
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: Brook, David E.
REGISTRATION NUMBER: 22,592
REFERENCE/DOCKET NUMBER: CPI95-08
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781) 861-6240
TELEFAX: (781) 861-9540

INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 3600 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 85..1128
FEATURE:
NAME/KEY: CDS
LOCATION: 1139..3559
US-08-855-910-7

Query Match 3.2%; Score 35.6; DB 4; Length 3600;
Best Local Similarity 51.9%; Pred. No. 3;
Matches 80; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 708 AACCTGCAACATGTTATTGACGCTGTGGAAAGCAAGTACTATCCCGAGGCTATTA 767
DB 3223 AAACCTGCAACATGTTATTGACGCAAGCAAGCGGCTTACTTTTGAAGCATTTCTAA 3282
QY 768 AGTCACAGAGATGATCTACCGCAATTTATTAAGGTTCATGATGACATTGAAAG 827
DB 3283 ATTCACAGAGATTTCTCGATATTTGCTTTATTTAGTCATGAAACGTTACCAATCAAGA 3342
QY 828 CATTCCTGAAGAAATCAGAAAGATATGTTAAA 861
DB 3343 ACTAGTTAAACTATTTGATGATACGACGTTAAA 3376

RESULT 4

US-08-992-334-1
; Sequence 1, Application US/08992334
; Patent No. 5919678
; GENERAL INFORMATION:

APPLICANT: Grus, Alexandra
APPLICANT: Maguin, Emmanuelle
TITLE OF INVENTION: METHODS FOR USING A TEMPERATURE-SENSITIVE
TITLE OF INVENTION: PLASMID
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESS: Christie Parker & Hale, LLP
STREET: 350 West Colorado Boulevard, Suite 500
CITY: Pasadena
STATE: California
COUNTRY: United States
ZIP: 91105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/992,334
FILING DATE: 17-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,752
FILING DATE: 24-DEC-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/FR93/00248
FILING DATE: 12-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 992/03034
FILING DATE: 13-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Prout, D. Bruce
REGISTRATION NUMBER: 20958
REFERENCE/DOCKET NUMBER: C93:31779
TELECOMMUNICATION INFORMATION:
TELEPHONE: (626) 795-9900
TELEFAX: (626) 577-8800
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3792 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOHETICAL: NO
ANTI-SENSE: YES
IMMEDIATE SOURCE:
CLONE: PG+host4
8-992-334-1

Query Match 3.2%; Score 35.6; DB 2; Length 3792;
Best Local Similarity 47.3%; Pred. No. 3;
Matches 107; Conservative 0; Mismatches 119; Indels 0; Gaps 0;
QY 820 ATTGAAGCATTCCTGAGAAATCAGAAAAGATATGTTAAATTGATTTACTTTGTTG 879
DB 2229 ATTGCACGAATCTCTGTAACAATAGAAAGCTTAGAACAAGTTAAAGCAAAATTTGGG 2288
QY 880 TATCGTCATCTGGTCAAGAAATTCATTTGGGTGCAACAAGATTTCTGCTAGTCCAGAA 939
DB 2289 AATAGTTCACTGCTCATGTTGAGATACCTGATTATATCAAAAGTTTCATGTGAATATTG 2348
QY 940 TATCAGTAAGCCAGCACTTAATTCGTTAAGTTTATTAATCCACCTTAACAAGGTGAT 999
DB 2349 ACTCATGAATCAAGACGCTATGCTAAGAAATTAATATTAACAAAAAGATATT 2408
QY 1000 ATCCATCATATGATATTAAGTTTTCAGCAACGGCTTAGATGATG 1045
DB 2409 TTGAACATTATGATTTTGTATTTAGACCGCTATATAACACTTGATG 2454

RESULT 5
US-08-302-752-1

Sequence 1, Application US/08302752
Patent No. 6025190
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: THERMOSENSIBLE PLASMID
NUMBER OF SEQUENCES: 3
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/302,752
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 9203034
FILING DATE: 13-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO FR/93/00248
FILING DATE: 12-MAR-1993
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3792 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-302-752-1

Query Match 3.2%; Score 35.6; DB 3; Length 3792;
Best Local Similarity 47.3%; Pred. No. 3;
Matches 107; Conservative 0; Mismatches 119; Indels 0; Gaps 0;
QY 820 ATTGAAGCATTCCTGAGAAATCAGAAAAGATATGTTAAATTGATTTACTTTGTTG 879
DB 2229 ATTGCACGAATCTCTGTAACAATAGAAAGCTTAGAACAAGTTAAAGCAAAATTTGGG 2288
QY 880 TATCGTCATCTGGTCAAGAAATTCATTTGGGTGCAACAAGATTTCTGCTAGTCCAGAA 939
DB 2289 AATAGTTCACTGCTCATGTTGAGATACCTGATTATATCAAAAGTTTCATGTGAATATTG 2348
QY 940 TATCAGTAAGCCAGCACTTAATTCGTTAAGTTTATTAATCCACCTTAACAAGGTGAT 999
DB 2349 ACTCATGAATCAAGACGCTATGCTAAGAAATTAATATTAACAAAAAGATATT 2408
QY 1000 ATCCATCATATGATATTAAGTTTTCAGCAACGGCTTAGATGATG 1045
DB 2409 TTGAACATTATGATTTTGTATTTAGACCGCTATATAACACTTGATG 2454

RESULT 6
US-08-992-334-2
Sequence 2, Application US/08992334
Patent No. 5919678
GENERAL INFORMATION:
APPLICANT: Grus, Alexandra
TITLE OF INVENTION: METHODS FOR USING A TEMPERATURE-SENSITIVE
TITLE OF INVENTION: PLASMID
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESS: Christie Parker & Hale, LLP
STREET: 350 West Colorado Boulevard, Suite 500
CITY: Pasadena
STATE: California
COUNTRY: United States
ZIP: 91105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/992,334
; FILING DATE: 17-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,752
; FILING DATE: 24-DEC-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/FR93/00248
; FILING DATE: 12-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 92/03034
; FILING DATE: 13-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Prout, D. Bruce
; REGISTRATION NUMBER: 20958
; REFERENCE/DOCKET NUMBER: C93:31779
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (626) 795-9900
; TELEFAX: (626) 577-8800
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5234 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; US-08-992-334-2

Query Match      3.2%; Score 35.6; DB 2; Length 5234;
Best Local Similarity 47.3%; Pred. No. 3.5;
Matches 107; Conservative 0; Mismatches 119; Indels 0; Gaps 0;

QY      820  ATTGAAGCATTCCTGAGAAGATCAGAAAGATATGTTAAATGATTACTTTGTTG 879
      |||||
DB      3671  ATTGACGAATTCCTGTAACATAGAAAGCTTAGAACAAAGTAAAGCAAAATTCGGG 3730
      |||||

QY      880  TATCGTCATCTGCTCAAGAAATTCATATGGGTGCAACAAGATTTCTGCTAGTCAGAA 939
      |||||
DB      3731  AATAGTTCAGTTGCTCATGTTGAGATCTGATTATATCAAGGTTGATATGAAATATTG 3790
      |||||

QY      940  TATCATGAAGCCAGTTAAATTCGTTAAGTTATATCAACCTTACACGCGTGAT 999
      |||||
DB      3791  ACTCATGAATCAAGAGCGCTATGCTTAAGAAATTAACATATATACGACAAAAAGATATT 3850
      |||||

QY      1000  ATCCATCATATGATATTAAGTTTTCAGCAACGCGTTAGATGATG 1045
      |||||
      3851  TTGAACATTATGATTTTGTATTTGACCGCTATATTAACACTTGATG 3896

RESULT 7
US-08-302-752-2
; Sequence 2, Application US/08302752
; Patent No. 6025190
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: THERMOSENSIBLE PLASMID
; NUMBER OF SEQUENCES: 3
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25 (BPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/302,752
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 9203034
; FILING DATE: 13-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO FR/93/00248
; FILING DATE: 12-MAR-1993
```

```

; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5234 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-302-752-2

Query Match      3.2%; Score 35.6; DB 3; Length 5234;
Best Local Similarity 47.3%; Pred. No. 3.5;
Matches 107; Conservative 0; Mismatches 119; Indels 0; Gaps 0;

QY      820  ATTGAAGCATTCCTGAGAAGATCAGAAAGATATGTTAAATGATTACTTTGTTG 879
      |||||
DB      3671  ATTGACGAATTCCTGTAACATAGAAAGCTTAGAACAAAGTAAAGCAAAATTCGGG 3730
      |||||

QY      880  TATCGTCATCTGCTCAAGAAATTCATATGGGTGCAACAAGATTTCTGCTAGTCAGAA 939
      |||||
DB      3731  AATAGTTCAGTTGCTCATGTTGAGATCTGATTATATCAAGGTTGATATGAAATATTG 3790
      |||||

QY      940  TATCATGAAGCCAGTTAAATTCGTTAAGTTATATCAACCTTACACGCGTGAT 999
      |||||
DB      3791  ACTCATGAATCAAGAGCGCTATGCTTAAGAAATTAACATATATACGACAAAAAGATATT 3850
      |||||

QY      1000  ATCCATCATATGATATTAAGTTTTCAGCAACGCGTTAGATGATG 1045
      |||||
DB      3851  TTGAACATTATGATTTTGTATTTGACCGCTATATTAACACTTGATG 3896

RESULT 8
US-08-992-334-3
; Sequence 3, Application US/08992334
; Patent No. 5919678
; GENERAL INFORMATION:
; APPLICANT: Gruse, Alexandra
; TITLE OF INVENTION: METHODS FOR USING A TEMPERATURE-SENSITIVE
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Christie Parker & Hale, LLP
; STREET: 350 West Colorado Boulevard, Suite 500
; CITY: Pasadena
; STATE: California
; COUNTRY: United States
; ZIP: 91105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/992,334
; FILING DATE: 17-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,752
; FILING DATE: 24-DEC-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/FR93/00248
; FILING DATE: 12-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 92/03034
; FILING DATE: 13-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Prout, D. Bruce
; REGISTRATION NUMBER: 20958
; REFERENCE/DOCKET NUMBER: C93:31779
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (626) 795-9900
; TELEFAX: (626) 577-8800
; INFORMATION FOR SEQ ID NO: 3:
```

SEQUENCE CHARACTERISTICS:
LENGTH: 6722 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
US-08-992-334-3

Query Match 3.2%; Score 35.6; DB 2; Length 6722;
Best Local Similarity 47.3%; Pred. No. 3.8; Indels 0; Gaps 0;
Matches 107; Conservative 0; Mismatches 119; Indels 0; Gaps 0;

QY 820 ATTGAAGCATTCCTGGAAGAAATCGAAAAGATTAATGTTAAATGATATTACTTTGTTG 879
DB 5159 ATTGCACGAATCTCTTAACATTAAGAACCGTTAGAACACAGATTAAACGAAATTTGGGG 5218
QY 880 TATCGTGCACTGTGTCAGAAATTTCTATTGGGTGACAGAAAGTTTCTGCTAGTCAGAA 939
DB 5219 AATAGTTCAGTTGCTCATGTTGAGATTAATCTGATTTATCAAAAGTTCAATATGATATTTG 5278
DB 940 TATCATGAAGCCACAGTTAAATCTGTTAAGTTTAAATCCACCTTAACAGCGTAT 999
DB 5279 ACTCATGATCAAGACGCTATTCTAGAAATTAACATATATACGACAAAAGATATTT 5338
QY 1000 ATCCATCATATGAATATTAAGTTTTCAGCAACGCGTTAGATGATG 1045
DB 5339 TTGAACATTATGATTTGATTTGATTTGACCGCTATATACACTTGATG 5384

RESULT 9
US-08-302-752-3
Sequence 3, Application US/08302752
Patent No. 6025190
GENERAL INFORMATION:

APPLICANT:
TITLE OF INVENTION: THERMOSENSIBLE PLASMID
NUMBER OF SEQUENCES: 3
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/302,752
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 9203034
FILING DATE: 13-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO FR/93/00248
FILING DATE: 12-MAR-1993
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 6722 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-302-752-3

Query Match 3.2%; Score 35.6; DB 3; Length 6722;
Best Local Similarity 47.3%; Pred. No. 3.8; Indels 0; Gaps 0;
Matches 107; Conservative 0; Mismatches 119; Indels 0; Gaps 0;

QY 820 ATTGAAGCATTCCTGGAAGAAATCGAAAAGATTAATGTTAAATGATATTACTTTGTTG 879
DB 5159 ATTGCACGAATCTCTTAACATTAAGAACCGTTAGAACACAGATTAAACGAAATTTGGGG 5218
QY 880 TATCGTGCACTGTGTCAGAAATTTCTATTGGGTGACCAAGATTTCTGCTAGTCAGAA 939
DB 5219 AATAGTTCAGTTGCTCATGTTGAGATTAATCTGATTTATCAAAAGTTCAATATGATATTTG 5278

QY 940 TATCATGAAGCCACAGTTAAATCTGTTAAGTTTAAATCCACCTTAACAGCGTAT 999
DB 5279 ACTCATGATCAAGACGCTATTCTAGAAATTAACATATATACGACAAAAGATATTT 5338
QY 1000 ATCCATCATATGAATATTAAGTTTTCAGCAACGCGTTAGATGATG 1045
DB 5339 TTGAACATTATGATTTGATTTGATTTGACCGCTATATACACTTGATG 5384

RESULT 10
US-09-134-001C-650
Sequence 650, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:

APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 650
LENGTH: 1434
TYPE: DNA
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-650

Query Match 3.2%; Score 35.2; DB 4; Length 1434;
Best Local Similarity 47.7%; Pred. No. 2.6; Indels 0; Gaps 0;
Matches 103; Conservative 0; Mismatches 113; Indels 0; Gaps 0;

QY 774 AGCAGATAGTCTACCTGCCACATTATTAAGATGGTTCCAATGACATTGAAGCATTC 833
DB 552 AGCTTTATTTTAGGTTCTTCTCGTAGAAATATCAAAAACAAAAGACATTTGCT 611
QY 834 TGAAGAAATCGAAGAAAGATTAATGTTAAATGATATTACTTTGTTGATCGTCATCGG 893
DB 612 TGATTAATTTCTGATGTTATTTCTACTTTCGCTTGGTTGTTAAATTTATTCGCAATTTAG 671
QY 894 TCAAGAAATCTATTGGGTGGAACAAGATTTCTGCTAGTCCAGATATCATGAAGCCAC 953
DB 672 TAGCGTAGACGTTGGTATACCTCTCTACCTGTCAATGTGACATTTGTACTGGTAT 731
QY 954 AGTAAATTCGTTAAGTTTATTAATTCACACCTTAA 989
DB 732 AGCATTATCATTAATTTTACAGCGCTCAGCTTAA 767

RESULT 11
US-09-134-001C-461

Sequence 461, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 461
LENGTH: 1044
TYPE: DNA
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-461

Query Match 3.1%; Score 34.8; DB 4; Length 1044;
Best Local Similarity 45.6%; Pred. No. 3;
Matches 123; Conservative 0; Mismatches 147; Indels 0; Gaps 0;

499 CCGCAACATCTATCATTTGATTTGGGCTGCTACAGAGGGCTCAACACTA 558
463 CTTAAGAGAGTGAAGAGTATTTTCAAGCTTCAAGCTTCAAGCTTCAAG 522
559 ATCCAGTTCGCAACATCAATGCTTATTAAGATTGTAACCTGTTCTTCAAA 618
523 GCGCAATTCGCAAGCTTAAAGTTGAGAGTGGTTGATAGCTGGTGAAGTAA 592
619 CATTGAAGAGCTTTTAAAGTTTATGCTGCTGATGATCTTTTGACTATCA 678
583 GTGAACATCTTAAAGAGTAACTCTGTTGATGCTGCTGATGCTTCAAAAG 642
679 GTTATGAGAGATCAATGGAAGTATCAACCTGCAAGATGTTATGAGCGCT 738
643 TTCCCTGAGAGCTTAAAGAGAGGCTTCAAGCTTATGATGCTTCAAAAT 702
739 AGCAAGATAGTATCCCGAGGCTTATAAA 768
703 GGTGATATATTTGGCGGATGAAGCTTCAA 732

RESULT 12.

US-08/622
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 1347 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
FEATURE:
NAME/KEY: CDS
LOCATION: 53..1168
SEQUENCE DESCRIPTION: SEQ ID NO: 11:
US-08/622, 679D-11

Query Match 3.1%; Score 34.8; DB 3; Length 1347;
Best Local Similarity 48.5%; Pred. No. 3.3;
Matches 96; Conservative 0; Mismatches 102; Indels 0; Gaps 0;

680 TTATTGAGAGATCAATGCAATGCAATCAACCTGCAACATGTTATGAGCT 739
978 TTCTGGGCAATTCCTCGAGCAACAGAACCAATATTTTGGGATGATTTCA 1037
740 GCGAAGATAGTATCCCGAGGCTTATTAAGTCAAGAGATAGTCTACCTGC 799
1038 CCGTGAAGTATGCAATGAGATATTTCAAGATTTCTGCGTTATCTCTCA 1097
800 TAGAAGTGTTCGAATGACCATGGAAGCATCTCTGAAGAAATCAGAAAAG 859
1098 TTATGCTGATCTCTTTCGATGGGCGCTTAACAGCAAAAAGTCAATTA 1157
860 AAATGATATTAATCTTGT 877
1158 AGATCAATGACTGTAT 1175

RESULT 13

US-09-165-922A-11
Sequence 11, Application US/09165922A
Patent No. 6348641
GENERAL INFORMATION:
APPLICANT: STILES, JOHN I.
MOISTYAD, ISTEFO
NEUPANE, KABI R.
TITLE OF INVENTION: PURIFIED PROTEINS,
RECOMBINANT DNA SEQUENCES AND PROCESSES FOR
PRODUCING CAFFEINE FREE BEVERAGES

NUMBER OF SEQUENCES: 11

CORRESPONDENCE ADDRESS:
ADDRESSEE: JONES, DAY, REAVIS & POGUE
STREET: NORTH POINT, 901 LAKESIDE AVENUE
CITY: CLEVELAND
STATE: OHIO
COUNTRY: USA
ZIP: 44114

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb
storage

COMPUTER: IBM PC compatible

OPERATING SYSTEM: MS-DOS v. 5.1

SOFTWARE: Wordperfect for Windows v. 6.1

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/165,922A

Filing DATE: 02-Oct-1998

CLASSIFICATION: 800

ATTORNEY/AGENT INFORMATION:
NAME: GRIFFITH, CALVIN P.

REGISTRATION NUMBER: 34,831

REFERENCE/DOCKET NUMBER: 26503660003

TELEPHONE: (216) 586-7050

TELEFAX: (216) 579-0212

INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 1347 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
FEATURE:
NAME/KEY: CDS
LOCATION: 53..1168
SEQUENCE DESCRIPTION: SEQ ID NO: 11:
US-09-165-922A-11

Query Match 3.1%; Score 34.8; DB 4; Length 1347;
Best Local Similarity 48.5%; Pred. No. 3.3;
Matches 96; Conservative 0; Mismatches 102; Indels 0; Gaps 0;

680 TTATTGAGAGATCAATGCAATGCAATCAACCTGCAACATGTTATGAGCT 739
978 TTCTGGGCAATTCCTCGAGCAACAGAACCAATATTTTGGGATGATTTCA 1037
740 GCGAAGATAGTATCCCGAGGCTTATTAAGTCAAGAGATAGTCTACCTGC 799
1038 CCGTGAAGTATGCAATGAGATATTTCAAGATTTCTGCGTTATCTCTCA 1097
800 TAGAAGTGTTCGAATGACCATGGAAGCATCTCTGAAGAAATCAGAAAAG 859
1098 TTATGCTGATCTCTTTCGATGGGCGCTTAACAGCAAAAAGTCAATTA 1157
860 AAATGATATTAATCTTGT 877
1158 AGATCAATGACTGTAT 1175

RESULT 14

US-09-173-053-3
Sequence 3, Application US/09173053
Patent No. 6451769
GENERAL INFORMATION:
APPLICANT: HUBNER, Robert C.
APPLICANT: NORMAN, Jon A.
APPLICANT: LIANG, Xiaowu
APPLICANT: CARNER, Kristin R.
APPLICANT: BARBOUR, Alan G.
APPLICANT: LUKE, Catherine J.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR ADMINISTERING BORRELIA DNA
FILE REFERENCE: 454312-2440.1
CURRENT APPLICATION NUMBER: US/09/173,053

/ CURRENT FILING DATE: 1998-10-15
 / PRIOR APPLICATION NUMBER: 08/663,998
 / PRIOR FILING DATE: 1996-06-14
 / NUMBER OF SEQ ID NOS: 18
 / SOFTWARE: Patent In Ver. 2.1
 / SEQ ID NO 3
 / LENGTH: 4016
 / TYPE: DNA
 / ORGANISM: Borrelia burgdorferi
 / FEATURE:
 / NAME/KEY: misc.feature
 / LOCATION: (6)..(29)
 / OTHER INFORMATION: N stands for A or G or C or T
 / NAME/KEY: misc.feature
 / LOCATION: (1078)..(1085)
 / OTHER INFORMATION: N stands for A or G or C or T
 / US-09-173-053-3

Query Match 3.1%; Score 34.8; DB 4; Length 4016;
 Best Local Similarity 47.0%; Pred. No. 5.2;
 Matches 142; Conservative 0; Mismatches 157; Indels 3; Gaps 1;

```

QY 803 AAGGTGTCATGACCATGGAAGCATTCCTGAAGAAATCAGAAAAGATTAATGTTAAA 862
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 36 AATTGAACCAACTTAATTAATAACCAACTTAATTAATTAATTAATTAATTAATTTT 95
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 863 TTGATATTAATCTTTGTTGATCGTCATCGTCAGAAATTTGATTTGGTGAACAAGT 922
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 96 TTCAATTTTCTATTGTTATGTTGTTATGTTATTAATTAATTAATTAATTAATTAAT 155
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 923 TTCCTGCTAGTCAGAAATATCATGAAAGCAGCATTAATTCGTTAATTAATTAATCCAC 982
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 156 TTAATTAATAAGAGAAATTAATTAAGAAATTAATTAATTAATTAATTAATTAATTAAT 215
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 983 ACCTTAACAGCGGTGATTCATCATATGAAATTAATTAATTTTCAGCAACGCTTAGATG 1042
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 216 CCTTAATGATGATGAAGCAAAATGTTAGCAGCTTGACAGAAAACACGTTTCAGTAG 275
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1043 ATGTCCCGCTCTCACTGAG--GTTAATAAGAGGTAAACAAATGTTAATGATAG 1099
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 276 ATTTGCTGTGTAAGAAAGTTCTTTGTAAGCAAGAAAACAAAGCGCAAGTAG 335
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1100 TT 1101
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 336 AT 337
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  
```

/ Patent No. 6090620
 / Sequence 209, Application US/08781891

/ GENERAL INFORMATION:
 / APPLICANT: Fu, Ying-Hui.
 / APPLICANT: Yu, Chang-En.
 / APPLICANT: Oshima, Junko.
 / APPLICANT: Mulligan, John T.
 / APPLICANT: Schellenberg, Gerald D.
 / TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO
 / NUMBER OF SEQUENCES: 209
 / CORRESPONDENCE ADDRESS:
 / ADDRESSEE: SEED and BERRY LLP
 / STREET: 6300 Columbia Center, 701 Fifth Avenue
 / CITY: Seattle
 / STATE: Washington
 / COUNTRY: USA
 / ZIP: 98104-7092

/ COMPUTER READABLE FORM:
 / MEDIUM TYPE: Floppy disk
 / COMPUTER: IBM PC compatible
 / OPERATING SYSTEM: PC-DOS/MS-DOS
 / SOFTWARE: Patent In Release #1.0, Version #1.30
 / CURRENT APPLICATION DATA:

/ APPLICATION NUMBER: US/08/781,891
 / FILING DATE: 27-DEC-1996
 / CLASSIFICATION: 800
 / ATTORNEY/AGENT INFORMATION:
 / NAME: No. 6090620 Leburg Ph.D., Carol
 / REGISTRATION NUMBER: 39,317
 / REFERENCE/DOCKET NUMBER: 240052.419
 / TELECOMMUNICATION INFORMATION:
 / TELEPHONE: (206) 622-4900
 / TELEFAX: (206) 682-6031
 / INFORMATION FOR SEQ ID NO: 209:
 / SEQUENCE CHARACTERISTICS:
 / LENGTH: 51259 base pairs
 / TYPE: nucleic acid
 / STRANDEDNESS: single
 / TOPOLOGY: linear
 / US-08-781-891-209

Query Match 3.1%; Score 34.8; DB 3; Length 51259;
 Best Local Similarity 49.2%; Pred. No. 15;
 Matches 124; Conservative 0; Mismatches 122; Indels 6; Gaps 1;

```

QY 187 TGTGACATTTGCTGTACAGTTGTCAAACTTGACCAAAATGCTAGTACTGACTTGAAGTT 246
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1440 TGGGATGTTATTTGTACAGTAGTGTGACATTAATTAATTAATTAATTAATTAAT 14459
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 247 GGAGATACCGTTTCGTTTTCATCGGTGCTTCCCAACAGATCCTAATAATGTTGCA 306
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 14460 AATGTTAATTTTATTTTAAAGTAGTCCGGGTCTAATAACAGAAATGACAGATACCTCA 14519
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 307 TTTGCTGAATATGACAGGTTTATCCACTTTGTTTAA-----CAGAGTAACCTTAAT 360
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 14520 TGGGATGCTTCCAGGGGTACTGTGATTTGCTTTTAACCTTGGAAATGACCTGTAAT 14579
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 361 CACTCACTGCTGATGAATTTCTGAAGCCCTGTGAAGAACTTGCAATTCGTCATCA 420
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 14580 GGCAGATGCTTAATAATGAATCTCTACAGACCTTGAAGACCCCTGAACCTTTGCAATCA 14639
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 421 TTGCAAGTTTG 432
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 14640 GAGTGAATTTTG 14651
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  
```

Search completed: June 17, 2003, 09:23:14
 Job time : 68 secs

THIS PAGE BLANK (USPTO)

Result No.	Score	Query Match	Length	DB	ID	Description
1	1113	100.0	1113	9	US-10-081-644-1	Sequence 1, Appl1
2	502.2	45.1	509	9	US-10-081-644-14	Sequence 14, Appl1
3	359	32.3	1134	9	US-10-081-644-5	Sequence 5, Appl1
4	357.8	32.1	1145	9	US-10-081-644-3	Sequence 3, Appl1
5	345.2	31.0	1122	9	US-10-081-644-7	Sequence 7, Appl1
6	45.4	4.1	23439	7	US-08-781-986A-38	Sequence 36, Appl1
7	40.2	3.6	439	10	US-09-864-761-20174	Sequence 20174, A
8	40	3.6	1236	9	US-10-077-584-3	Sequence 3, Appl1
9	38.8	3.5	392	10	US-09-878-574-1620	Sequence 1620, Appl1
10	37.6	3.4	640681	10	US-09-790-998-1	Sequence 1, Appl1
11	37.2	3.3	7772	9	US-10-198-846-14020	Sequence 14020, A
12	37	3.3	8772	10	US-09-788-711A-3	Sequence 3, Appl1
13	37	3.3	8871	10	US-09-788-711A-1	Sequence 1, Appl1
14	37	3.3	198285	10	US-09-880-107-3814	Sequence 3814, Appl1
15	36.8	3.3	441	10	US-09-868-761-3403	Sequence 3403, Appl1
16	36.8	3.3	3150	9	US-09-938-842A-243	Sequence 243, Appl1
17	36.6	3.3	969	10	US-09-875-242-6606	Sequence 6606, Appl1
18	36.6	3.3	6283	10	US-09-070-927A-160	Sequence 160, Appl1
19	36	3.2	312	10	US-09-815-242-7011	Sequence 7011, Appl1

C	20	36	3.2	1830121	9	US-10-339-860-1	Sequence 1, App1
C	21	35.6	3.2	429	10	US-09-815-242-7144	Sequence 7144, App1
C	22	35.6	3.2	2421	10	US-09-815-242-3860	Sequence 3860, App1
C	23	35.6	3.2	2424	10	US-09-815-242-6622	Sequence 6622, App1
C	24	35.6	3.2	10929	9	US-10-032-353-7	Sequence 7, App1
C	25	35.4	3.2	520	10	US-09-864-761-8552	Sequence 8552, App1
C	26	35.4	3.2	1803	10	US-09-864-756-812	Sequence 812, App1
C	27	35.4	3.2	1959	10	US-09-864-761-4012	Sequence 4012, App1
C	28	35.2	3.2	405	10	US-09-960-352-13651	Sequence 13651, App1
C	29	35.2	3.2	2000	9	US-09-938-842A-3404	Sequence 3404, App1
C	30	35.2	3.2	640681	10	US-09-720-886-1	Sequence 1, App1
C	31	35	3.1	1882	10	US-09-070-527A-842	Sequence 842, App1
C	32	35	3.1	5513	10	US-09-800-065-3	Sequence 3, App1
C	33	34.8	3.1	360	9	US-09-736-457-1408	Sequence 1408, App1
C	34	34.8	3.1	360	9	US-09-902-941-1408	Sequence 1408, App1
C	35	34.8	3.1	360	9	US-09-849-626-1408	Sequence 1408, App1
C	36	34.8	3.1	360	9	US-10-017-754-1408	Sequence 1408, App1
C	37	34.8	3.1	456	10	US-09-864-761-4249	Sequence 1249, App1
C	38	34.8	3.1	3931	9	US-10-006-780-1	Sequence 1, App1
C	39	34.6	3.1	477	10	US-09-864-761-5436	Sequence 5436, App1
C	40	34.6	3.1	503	9	US-10-091-572-878	Sequence 878, App1
C	41	34.6	3.1	503	9	US-09-764-891-9331	Sequence 9331, App1
C	42	34.6	3.1	1080	9	US-09-938-842A-231	Sequence 231, App1
C	43	34.6	3.1	1283	10	US-09-887-576-601	Sequence 601, App1
C	44	34.4	3.1	418	9	US-09-991-936-213	Sequence 213, App1
C	45	34.4	3.1	2429	9	US-10-149-819-88	Sequence 28, App1

ALIGNMENTS

RESULT 1

US-10-081-644-1
; Sequence 1, Application US/10081644

; GENERAL INFORMATION:

APPLICANT: Yamamoto, Hiroaki

APPLICANT: KIMOTO, NO. US20020192782A11h1fo

TITLE OF INVENTION: NOVEL ENONE REDUCTASES, METHODS FOR PRODUCING SAME AND METHODS FOR SEPARATING SAME

TITLE OF INVENTION: CARBON-CARBON DOUBLE BOND OF AN ALPHA. BETA-UNSATURATED KETONE

1. TITLE OF INVENTION: USING THE REDUCTASES

FILE REFERENCE: 06501-1

; CURRENT APPLICATION NUMBER: US/10/081,644

CURRENT FILING DATE: 2002-02-21

PRIOR APPLICATION NUMBER: JP 2001-49363

PRIOK FILING DATE: 2001-11-01

```

; NUMBER OF SEQ ID NOS: 23
: SOFTWARE: FastSeq for Windows Version 4.0

```

SEO ID NO 1

LENGTH: 1113

TYPE: DNA

ORGANISM: *Kluyveromyces fragilis*

```

; FEATURE:
NAME /xxx/ and

```

NAME/KEY:	CDS
LOCATION: (1)	(11110)

LOCATION: (

C
 C
 F
 C
 C
 C
 F
 F
 F
 F

Query Match

Best Local Similarity

Matches 1113; Conservat

RECEIVED

1 AIGI CAGI I CCH
QY

1 ATCTCAGTTCTA

100

61 ACAGATGTCCTCA

Db 61 ACAGATGTCCTCA

2000

121 GC1GGTAACCCCA

121 GCTGCTTAACTCTA

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

PRIOR FILING DATE: 2001-02-23
 NUMBER OF SEQ ID NOS: 29
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 5
 LENGTH: 1134
 TYPE: DNA
 ORGANISM: Saccharomyces cerevisiae
 FEATURE:
 NAME/KEY: CDS
 LOCATION: (1)...(1133)
 US-10-081-644-5

Query Match 32.3%; Score 359; DB 9; Length 1134;
 Best Local Similarity 58.9%; Pred. No. 4.6e-84;
 Matches 658; Conservative 0; Mismatches 450; Indels 9; Gaps 2;

```

QY 5 CAGTTCACCACTCAAAAAGCCGTCATATTGAAGTGACAAAGCTGTGTTAAACAG 64
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 14 CAATTCAGAAACATGAAAGCGTCGTCATTGAAAGACGCTAAAGCGGTTGTTAAAGAG 73
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
65 ATGTCGCTCCAGATTAAAGAGGGTACAGCCCTGGTGAAGGTGAGCGTGTGCTG 124
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 74 GCATTCCTCAATCTTAATTGAGAAAGATTCGTATGATTAAGACACTCGCTGTGCTG 133
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 125 GTAACCCCACTGATTGGAAGCATATTGCTTATAGATTGGTCCAGAGGTTCAATTCTAG 184
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 134 GTAACCCCACTGATTGGAAGCATATTGCTTATAGATTGGTCCAGAGGTTCAATTCTAG 193
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 185 GATGTCATGCTGCTGACAGTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 241
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 194 GATGTCATGCTGCTGACAGTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 253
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 242 AGGTGGAGATCCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTC 301
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 254 CTATGCGTATATATATATATATATATATATATATATATATATATATATATATATATAT 313
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 302 GTGCAATTTGCTGAATATGCGAGGTTTATCCACTTTTGTGTTACAA-----GAGTAACT 355
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 314 GTGCTTTGCTGAATATGCGAGGTTTATCCACTTTTGTGTTACAA-----GAGTAACT 373
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 356 TAACTACTCAACTGCTGATGAAATTTTGAAGCCCTGCTGGAAGAACTTGAATCTGCTG 415
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 374 TCAATTTTGGGTGAGATGTTTACCTGCGGCCCTGCTGAGGTCTTGTGAAGGTGAG 433
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 416 CATCATGTCATTTGCTGACAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 475
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 434 CCATATCCCAAGTGTACAGACACAGCGGCTTGTGTGACTTAATTAATGAGGCTGTG 493
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 476 AAATGGAATGCAACCATCTACCCCGCAACATATCATCTCATTTATTTGAGGTGCTG 535
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 494 ACCTGAAGTGGAGGCCATTAACCCCAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 553
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 536 CTACAGCAAGTGGTCAACAACTAATTCAGATTGCGCAACATATCAATGCTTATCTAAGA 595
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 554 CAATGCAAGTGGTCAACAACTAATTCAGATTGCGCAACATATCAATGCTTATCTAAGA 613
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 556 TTGTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 655
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 614 TCATGTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 673
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 656 TCTTTGACATCATATGAGAGCGCTTATTTGAGCAGATCAATGCAAGTATTCGAACCTG 715
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 674 TATTTGATTTATATATATATGAGAGCGCTTATTTGAGCAGATCAATGCAAGTATTCGA 733
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 716 AACATGTTATGACGCTGTGGAAGGAAAGATATATCCCGAGGCTTATTAAGTCAAG 775
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 734 CGTATTTAGTCACTGTGTCGGAATCAAGATACGCTTCAACAAGTGTCAATATGCGG 793
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 776 CAGATGTCATCTGTCACATATTTAGAGTGTTCATGACATGCAATGAAGATTTCTG 835
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 794 CCGATTAACAGAGATGCTACAAATTTGGAATTTAAATTTGACAGAGAAAGCTCAAAA 853
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 836 AAGAAATCAGAAAGATATGTTAAATGATATATCTTGTGTATCGTCACTGCTGTC 895
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  
```

```

DB 854 AAGAGACAGAGACAAACCTACTATGACATATTAAGCTATATTCATAGGTGCCC 913
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 856 AAGAAATCTATTTGGTGTCCACAAAGATTTCTGCTAGTCCGAATATGAAAGCCACAG 955
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 914 ATGAAGTACCAATTTGGAACATTTACTTACACAGCCGACTCAGAACTGAGAAAGCTGCA 973
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 956 TTAATTTGTTAAGTTTAAATATCCACCTTAACACGAGTATATCCATCATATGATA 1015
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 974 TAAATTTTATCAATTTCAATTCGAATTAAGATTAAGTGAACAAATTTGCCATATTCAG 1033
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1016 TTAAGTTTTCAGCAAGCGCTTAAATGATGTCCTCAGCTTCACATGAAGTATTAAGAAG 1075
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1034 TAAAGTCTATTAAGAACCGGCTTGTGATGTTCTCATATCTTAAAGACATCAATATG 1093
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1076 GTAACCAACAAATTTTAAATGATGTCCTGAGTTATA 1112
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1094 GTAACCACTGTGTGAACAACTGCTTGTGCTGATTTAA 1130
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  
```

RESULT 4

US-10-081-644-3
 Sequence 3, Application US/10081644
 Publication No. US20020192782A1
 GENERAL INFORMATION:
 APPLICANT: Yamamoto, Hiroaki
 TITLE OF INVENTION: NOVEL ENONE REDUCTASES, METHODS FOR
 TITLE OF INVENTION: PRODUCING SAME, AND METHODS FOR SELECTIVELY REDUCING A
 TITLE OF INVENTION: CARBON-CARBON DOUBLE BOND OF AN ALPHA, BETA-UNSATURATED KETONE
 FILE REFERENCE: 06501-100001
 CURRENT APPLICATION NUMBER: US/10/081,644
 PRIOR FILING DATE: 2002-02-21
 PRIOR APPLICATION NUMBER: JP 2001-49363
 NUMBER OF SEQ ID NOS: 29
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 3
 LENGTH: 1145
 TYPE: DNA
 ORGANISM: Saccharomyces cerevisiae
 FEATURE:
 NAME/KEY: CDS
 LOCATION: (6)...(1133)
 US-10-081-644-3

Query Match 32.1%; Score 357.8; DB 9; Length 1145;
 Best Local Similarity 58.8%; Pred. No. 9.6e-84;
 Matches 658; Conservative 0; Mismatches 452; Indels 9; Gaps 2;

```

QY 4 TCAGTTCACCACTCAAAAAGCCGTCATATTGAAGTGACAAAGCTGTGTTAAACA 63
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 15 TCGATTTCCAGAAACCATGAAGCCGTTGCTATTTGAAATGCGAAGCGTGTGCTCAACAG 74
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 64 GATGTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 123
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 75 GACATTTCAATTTCTGAAATTTGAAGAGATTTGTTCTTAATTAACATGTCGCGGTGCT 134
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 124 GGTAAACCACTGATTTGAAGATATTTGCTTATTAAGTTGCTGCAAGTTCAATTTCTA 183
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 135 GGTAAACCACTGATTTGAAGATATTTGCTTATTAAGTTGCTGCAAGTTCAATTTCTA 194
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 184 GATGTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 195 GGTGTCATGAGCGCGGCAATCTTAAGTGTGAGGCGCAAAATGTTGATGCTGCAAGCTTT 254
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 241 AAGTTGAGATATACCGTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 300
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 255 GCCATTTGATTTACATTTATGAGGTTATTTACAGGTCCTTCAAGTATTCCTCAAC 314
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 301 GGTGATTTGCTGATATATGCAAGGTTTATCCACTTGTGTTTACAA-----GAGTAA 354
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  
```

```

Db 315 GGTGCTTTGCTGAGTACTGTCCTTTCATCCGAGACGTGCTTATTAACCAAGCCAGAG 374
Qy 355 TTAACCTACTCACTGCTGATGAAATTTCTGAGGCGCTTGAAGAACTTGAATCTGCT 414
Db 375 TTTAGATGTCGGGTAAGACAGCTACCAAGAGCCCGCTTAATCTTTAGAGGGGCA 434
Qy 415 GCATCATGCTGCTTCTGTTGCAACATGCTGCTGTTAGTTAGTTGTCATCACTTGGCTCA 474
Db 435 GTATCCCTCCCATCTCATTCATGACCAAGCGCTGATGATCTTACATAGTCTTGGCTTG 494
Qy 475 AAAATGGAATGGAGCCCATCTACCCGCAACATCATCATCTATTTAGTTGGGGTGT 534
Db 495 GACATGATGAGAGCCCTCCAAAGGCAAGAGATCAACCATCTTATTTGGGGTGT 554
Qy 535 GCTACAGACAGTGGGTCAACCACTTAATCCAGTTGCCAAACATATCAATGCTTACTAG 594
Db 555 GCCACGTGCTTGGCCAGATGCTTATTCATTTGGCAAAAACATTAACGCTTTCAGAG 614
Qy 595 ATTGTAATGCTGCTTCTTAAAGCATGAAAGCTTTTAAAGTCTTATGCTGCTGATGAT 654
Db 615 ATCATGCTGCTGCTTCTTCTTAAACATGAAATTTGTTGAAGATACGCTGCAATGAA 674
Qy 655 GTCTTGAATGATATGATGAGGCTTATGAGAGATCAATGAGATTCACAACTG 714
Db 675 CTTTGTGATCAACAGATGCTGAGTATGAAACAGTAAAGAAAGTAAACAACT 734
Qy 715 CAACATGTTATGACGCTGAGGAGGAGAGATAGTATCCCGAGGCTTAAAGTCA 774
Db 735 CTTTGTGATCAACAGATGCTGAGTATGAAACAGTAAAGAAAGTAAACAACT 794
Qy 775 GCATGATGCTTACCTGCAATTTAGAGTGTCCAAATGACATTTGAAGCACTTCT 834
Db 795 GCTGATGATCTTACAGCTGAGTGTGCTTCAATGACCTTTTAAACGAAAGATATCAAG 854
Qy 835 GAGAAATGAGAAAGTATGTTAAATGATATTAATCTTGTATGCTGATGCTGT 894
Db 855 GAGAAAGACAGAGGCAAGGCTGATGATGAAAGAAAGCTTCTATATTTGATGAGAGT 914
Qy 895 CAAGAAATCTATTTGGGTGCAACAGATTTCTGCTAGTCCAGATATCATGAGCCACA 954
Db 915 AAGAGCTGCTCATTTGGCAAGTTACTTTGCCAGAGACCTGATATCAAGAGAGCCGC 974
Qy 955 GTTAAATGCTTAAATTTAAATTCACACCTTAAACAGGTGATTCATATGAT 1014
Db 975 ATTAATTTATTTAATGATCATCAATCCAAATCAATGATGTAATTCACACATCCCA 1034
Qy 1015 ATTAAGTTTGAAGAGGCTTAAAGATGATGCTCCAGCTCACTGAAGTATTAAGAA 1074
Db 1035 GTTAAGTTTGAAGAGGCTTAAAGATGATGATGCTCCAGCTTCACTGATATTAAGAC 1094
Qy 1075 GGTAAACAAATATGTTAATGTTGCCAGGTTATA 1113
Db 1095 GGAAGAAATCTGCGCAAAAGTTGGTTGCGCTTGA 1133

```

RESULT 5

```

US-10-081-644-7
; Sequence 7, Application US/10081644
; Publication No. US20020192782A1
; GENERAL INFORMATION:
; APPLICANT: Yamamoto, Hiroaki
; TITLE OF INVENTION: NOVEL ENONE REDUCTASES, METHODS FOR
; TITLE OF INVENTION: PRODUCING SAME, AND METHODS FOR SELECTIVELY REDUCING A
; TITLE OF INVENTION: CARBON-CARBON DOUBLE BOND OF AN ALPHA, BETA-UNSATURATED KETONE
; FILE REFERENCE: 06501-100001
; CURRENT APPLICATION NUMBER: US/10/081,644
; PRIOR APPLICATION NUMBER: JP 2001-49363
; PRIOR FILING DATE: 2001-02-23
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: FastSeq for Windows Version 4.0

```

```

; SEQ ID NO 7
; LENGTH: 1122
; TYPE: DNA
; ORGANISM: Saccharomyces cerevisiae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (7)...(1110)
US-10-081-644-7
Query Match 31.0%; Score 345.2; DB 9; Length 1122;
Best Local Similarity 58.5%; Pred. No. 1.9e-80;
Matches 643; Conservative 0; Mismatches 448; Indels 9; Gaps 2;

```

QY 913 GCAACAGATTTCTGCTAGTCCAGATATCATGAAGCCACAGTTAAATTCGTTAGTT 972
Db 910 GGCATATCTTCCCTGCTGACCCAGAGAGAGCTGCCAGAAATTCGCAAGTTCC 969
QY 973 ATTAATCCACACCTTTACCAAGGTGATATCCATCATATTAATTTAAAGTTTCACAG 1032
Db 970 ATCAATCCAAATTAAGTATGATGGCAAAATTCACCATATTCAGCAAGGCTCATATAGAAC 1029
QY 1033 GCGTATGATGATGCTCCAGCTCTCACTGAAGGTATTAAGGTAAGAAACAAATGTT 1092
Db 1030 GGGCTTTACGATGTTCTCTGTTATCTGGAAGACATTAAATGGTAAAGTCTGTGTGA 1089
QY 1093 AAGTATGTTGCCAGGTTATA 1112
Db 1090 AAACCTGTTGCCGTTATAA 1109

RESULT 6

US-08-781-986A-38/c

Sequence 38, Application US/08781986A
Publication No. US20030054436A1

GENERAL INFORMATION:

APPLICANT: Charles Kunsch
TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
NUMBER OF SEQUENCES: 5255
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/781,986A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Benson, Bob
REGISTRATION NUMBER: 30,446
REFERENCE/DOCKET NUMBER: PB248PP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 38:
SEQUENCE CHARACTERISTICS:
LENGTH: 23439 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-781-986A-38

Query Match 4.1%; Score 45.4; DB 7; Length 23439;
Best Local Similarity 45.4%; Pred. No. 0.6;
Matches 163; Conservative 0; Mismatches 196; Indels 0; Gaps 0;

QY 504 ACATATCTATCATATTTATTTGGGTGCTGACAGAGTGGTCAACAACTAATCCA 563
Db 16078 ACACAGCTTTTAACTACGTATGACAGTGGCAGAGTAAATGATTAATATGATTA 16019
QY 564 AGTGGCAACATGATGATGTTATCTAAGATGTTGCTTTCAAAAGCATGA 623
Db 16018 CATGCTGATTAATCAATTAATTAATTAATGACAGATGAACCTTAACACTTCC 15959
QY 624 AAAGCTTTAAAGTCTTATGCTGATGATGCTTTGATCATGATGACAGGCTTAT 683

Db 15958 ACCTGTTTCATTTGGCATTTGATGAGGATTTATGATTTCCCTTGAGAAAGCGAAC 15899
QY 684 TGAGCAGATCAATGAGATGATTCGAACCTGCAATGTTATGACGCTGNGGAAGCA 743
Db 15898 TATCATGATCATGTTTCCACATTTGACATTTGAAAGCGGTAAATAACGACATGATTT 15839
QY 744 AGATGATATCCCGAGGCTTAAAGTACAGACAGATAGTCTGACCATTAATTA 803
Db 15838 TGATGCTGAGCTTAATGAGAAAGCATCATCTATTATGTAAGTTGTTGATTTCTTAA 15779
QY 804 AGTGTTCATATGACCATTTGAAGCATTCCTGGAAGAAATCAGAAAGATATGTTAAA 862
Db 15778 CGCTATTAACAATGATGATATATATCATTTATTTAAAAAATCAATTTATCATCATGATA 15720

RESULT 7

US-09-864-761-20174/c

Sequence 20174, Application US/09864761
Patent No. US20020048763A1

GENERAL INFORMATION:

APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David K.
APPLICANT: Hanzel, David K.
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aomics-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
SEQ ID NO 20174
LENGTH: 439
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AL035419.9

; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 640681
; TYPE: DNA
; ORGANISM: Buchnera sp.
US-09-790-988-1

Query Match 3.3%; Score 37.6; DB 10; Length 640681;
Best Local Similarity 48.6%; Pred. No. 4,1e+02;
Matches 103; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

QY 716 AACATGTTATGACCTGTGGAGAGATAGTATCCCGAGCCTATAAGTCACAG 775
DB 547422 AATTGTTATGATGAGAACTTAACGAGACTATGATCCAGAAAGCTATTGACAG 547363
QY 776 CAGATAGTCTACCTGCCACATTATTAAGTGTCCAAATGACCAATGAAGCATTCTCTG 835
DB 547362 CAGTACTATTTTACAGAACTAGAAAGCATTTGTTGATTAAGATATTCGTGAAC 547303
QY 836 AAGAAATCGAAGAAATATGTTAAATGATTAATCTTGTGTATCGTCATCTGCTC 895
DB 547302 CTGAATTCAGAAAGAAAGAAAGCTGAATTTGAACCTAATTTGTTAGCTCAATGATGAT 547243
QY 896 AAGAAATCTATGCTGCTGCAACAGATTCTCT 927
DB 547242 TAGAACTTACAGTCTGCTCTGCTAATGCTT 547211

RESULT 11
US-10-198-846-14020/c
; Sequence 14020, Application US/10198846
; Publication No. US2003009974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; CURRENT FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; PRIOR FILING DATE: 2001-07-18
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14020
; LENGTH: 7772
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-198-846-14020

Query Match 3.3%; Score 37.2; DB 9; Length 7772;
Best Local Similarity 51.9%; Pred. No. 45;
Matches 84; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

QY 936 AGAATATCATGAGCCACAGTTAAATTCGTTAAGTTATTAATCCACACTTACACAGCG 995
DB 3446 AATATTCCTCGATATGATTTAAAAAATAAGTTATGAGACAACATATTAATTAAGT 3387
QY 996 TGATATCCATCATATGAATTAATTAAGTTTTCAGCAAGCCTTATGATGATGCTCCAGCTT 1055
DB 3386 CAATTTACAGTTAATTAATTAATTTCAATTTACAGCTACACACTATTAATTAACGAAATTA 3327
QY 1056 CACTGAAGTATTAAGAGATTAATAAACAATAATGTTAGTA 1097
DB 3326 AATTCAGTAAACAGAAAAAGAGGGAATAATAAAGCA 3285

RESULT 12
US-09-788-711A-3/c

; Sequence 3, Application US/09788711A
; Patent No. US20020058328A1
; GENERAL INFORMATION:
; APPLICANT: Tania Tamsin Testa
; TITLE OF INVENTION: NOVEL COMPOUNDS
; FILE REFERENCE: GP-30225
; CURRENT APPLICATION NUMBER: US/09/788,711A
; CURRENT FILING DATE: 2001-02-20
; PRIOR APPLICATION NUMBER: 0004196.2
; PRIOR FILING DATE: 2000-02-19
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 8772
; TYPE: DNA
; ORGANISM: HOMO SAPIENS
US-09-788-711A-3

Query Match 3.3%; Score 37; DB 10; Length 8772;
Best Local Similarity 58.7%; Pred. No. 55;
Matches 64; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 11 CAACCACTCAAAAAGCCGTCATCATTTGAAGTGACAAAGCTGTGTTAAACAGATGCT 70
DB 5456 CAGCTGCAGGAATAGCTGCCAGTCGTTGCTGCAATAGCTGTAGCAGACACGGGTTT 5397
QY 71 CAGTTCAGAAATTAAGAGGGGTACAGCCTTGTTGAAGTTGAGCTCT 119
DB 5396 GAGTCACAAAGGCTCAGGACGCTACAGCCTTGCTCCAGTTGATGCTCT 5348

RESULT 13
US-09-788-711A-1/c
; Sequence 1, Application US/09788711A
; Patent No. US20020058328A1
; GENERAL INFORMATION:
; APPLICANT: Tania Tamsin Testa
; TITLE OF INVENTION: NOVEL COMPOUNDS
; FILE REFERENCE: GP-30225
; CURRENT APPLICATION NUMBER: US/09/788,711A
; CURRENT FILING DATE: 2001-02-20
; PRIOR APPLICATION NUMBER: 0004196.2
; PRIOR FILING DATE: 2000-02-19
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 8871
; TYPE: DNA
; ORGANISM: HOMO SAPIENS
US-09-788-711A-1

Query Match 3.3%; Score 37; DB 10; Length 8871;
Best Local Similarity 58.7%; Pred. No. 55;
Matches 64; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 11 CAACCACTCAAAAAGCCGTCATCATTTGAAGTGACAAAGCTGTGTTAAACAGATGCT 70
DB 5456 CAGCTGCAGGAATAGCTGCCAGTCGTTGCTGCAATAGCTGTAGCAGACACGGGTTT 5397
QY 71 CAGTTCAGAAATTAAGAGGGGTACAGCCTTGTTGAAGTTGAGCTCT 119
DB 5396 GAGTCACAAAGGCTCAGGACGCTACAGCCTTGCTCCAGTTGATGCTCT 5348

RESULT 14
US-09-880-107-3814/c
; Sequence 3814, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.

```

1 TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
2
3 FILE REFERENCE: 44921-5028-WO
4
5 CURRENT APPLICATION NUMBER: US/09/880,107
6
7 PRIOR FILING DATE: 2001-06-14
8
9 PRIOR APPLICATION NUMBER: US 60/211,379
10
11 PRIOR FILING DATE: 2000-06-14
12
13 PRIOR APPLICATION NUMBER: US 60/237,054
14
15 PRIOR FILING DATE: 2000-10-02
16
17 NUMBER OF SEQ ID NOS: 3950
18
19 SOFTWARE: PatentIn Ver. 2.1
20
21 SEQ ID NO 3814
22
23 LENGTH: 198265
24
25 TYPE: DNA
26
27 ORGANISM: Homo sapiens
28
29 FEATURE:
30
31 OTHER INFORMATION: Genbank Accession No. US20020142981A1 X87344
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
9
```

Query Match	Score	DB	Length
Local Similarity	3.3%	10	198285
Best Local Similarity	49.2%	Pred. No. 3.2e+02	
Matches 97; Conservative	0	Mismatches 100	Indels 0; Gaps 0

Qy	829	ATTTCCTGAGAAATACGAAAGAAATATAGTAAATATGATATATTACTTTGTTGATGTCGA	888
Db	169231	ATTTCATTTTAAAGAAATATCTAAATATGTAATACCTTGACTTTTCCATCCACTTTGC	169172
Qy	889	TCTGCTCAGAAATTCCTATGTTGGGTGCACAAGAATTCCTCGTAGTCCAGATATCATGAA	948
Db	169171	TTTTCTTAGCGATATTATATGATATTTCAACTTTTTCACGTTAAATACAAATATCTTGG	169112
Qy	949	GCCACAGTTAAATTCGTTAGTTTAAATCCACACTTAACAAGGTATATCCATCAT	1008
Db	169111	GCTAGAGTTTCTTTGTTTTTTTTTTTAAATACAGGATTAACCTACAGAGAACTAAACAT	169052
Qy	1009	ATGAAATATTAAGTTTT	1025
Db	169051	AGTATATATATTAATAATT	169035

```

RESULT 15
US-09-864-761-3403/C
; Sequence 3403, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aecmca-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30

```

```

PRIORITY APPLICATION NUMBER: PCT/US01/00668
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00663
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00662
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00661
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00670
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: US 60/234,687
PRIORITY FILING DATE: 2000-09-21
PRIORITY APPLICATION NUMBER: US 09/608,408
PRIORITY FILING DATE: 2000-06-30
PRIORITY APPLICATION NUMBER: US 09/774,203
PRIORITY FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
SEQ ID NO 3403
LENGTH: 441
TYPE: DNA
ORGANISM: Homo sapiens.
FEATURES:
OTHER INFORMATION: MAP TO AL035419.9
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.3
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.2
OTHER INFORMATION: EXPRESSED IN HEPA, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.4
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN HEC100, SIGNAL = 1.5
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.6
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.5
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.5
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.2
US-09-864-761-3403

```

Query Match .	3.3%	Score 36.8;	DB 10;	Length 441;
Best Local Similarly	47.8%;	Pred. NO. 11;		
Matches 107; Conservative	0;	Mismatches 117;	Indels 0;	Gaps 0;

[illegible]

Search completed: June 17, 2003, 10:47:53
Job time : 201 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 06:21:51 ; Search time 1868 Seconds
(without alignments)
9649.666 Million cell updates/sec

Title: US-10-081-644-1
Perfect score: 1113
Sequence: 1 atgcagctccacacaccca.....agtagtcgcaggtatata 1113

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pin:*
21: em_gss_vit:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rtd:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
C 1	294.4	26.5	1011	17	CNS06N13	AL406189 T7 end of
C 2	182.4	16.4	886	17	CNS06N13	AL404058 T3 end of
C 3	181.2	16.3	881	17	CNS06N13	AL404057 T7 end of
C 4	154.6	13.9	1072	17	CNS06N13	AL404614 T7 end of
C 5	146.6	13.2	488	17	AZ931755	AZ931755 474.dhz89
C 6	137.2	12.3	1118	17	CNS06LWS	AL404715 T3 end of

C 7	132.6	11.9	434	17	AZ931754	AZ931754 474.dhz89
C 8	126.6	11.4	454	17	A0501781	A0501781 V14B1 mt
C 9	125.6	11.3	546	17	A0873738	A0873738 V7A3 mt
C 10	125	11.2	537	17	A0873797	A0873797 V7SE1 mt
C 11	119	11.1	529	17	A0502797	A0502797 V7SD10 mt
C 12	117.8	10.7	946	17	CNS06H8	AL398958 T7 end of
C 13	117.8	10.6	542	17	A0873736	A0873736 V7A41 mt
C 14	115.4	10.4	337	14	T38505	T38505 EST104002 S
C 15	108.4	9.7	540	17	AZ926593	AZ926593 476.dhz89
C 16	97.6	8.8	404	17	AZ926284	AZ926284 476.dhz89
C 17	96.4	8.7	353	17	A0501976	A0501976 V11A3 mt
C 18	90.2	8.1	549	17	A0876344	A0876344 V9B9 mt
C 19	79.6	7.2	522	17	AZ930225	AZ930225 474.dhz51
C 20	61.2	5.5	454	17	A0501614	A0501614 V17E2 mt
C 21	56.2	5.0	1146	17	CNS07B3	AL437653 T3 end of
C 22	52.2	4.7	556	14	B0143479	B0143479 fmlc.PK0
C 23	51.6	4.6	515	13	B1516433	B1516433 B160022A
C 24	51.6	4.6	619	13	B1505803	B1505803 B170024A
C 25	51.6	4.6	657	13	B1503620	B1503620 B170018A
C 26	51.6	4.6	657	13	B1506164	B1506164 B170014B
C 27	51.6	4.6	669	13	B1504612	B1504612 B170025B
C 28	51.6	4.6	706	13	B1508141	B1508141 B170011A
C 29	47.2	4.2	542	12	BE920514	BE920514 EST424283
C 30	47.2	4.2	620	12	BG096522	BG096522 EST461041
C 31	46.6	4.2	541	14	B0090513	B0090513 rc66d10.y
C 32	46.6	4.2	544	13	B1480612	B1480612 EST0065.H
C 33	46.2	4.2	925	17	BH155926	BH155926 ENTJ37TF
C 34	46	4.1	1101	17	CNS00FVE	AL071298 Drosophila
C 35	45.8	4.1	315	12	BE923966	C90538 C90538 Dict
C 36	44.8	4.0	389	12	BE923966	BE923966 EST427735
C 37	44.8	4.0	514	9	AJ273783	AJ273783 AJ273783
C 38	44.8	4.0	775	14	B0517113	B0517113 EST624528
C 39	44.2	4.0	416	12	BE924284	BE924284 EST428053
C 40	44.2	4.0	463	12	BE923969	BE923969 EST427738
C 41	44.2	4.0	624	12	BG521829	BG521829 15-14 Ste
C 42	44.2	4.0	694	17	AZ522107	AZ522107 202PbC01
C 43	43.4	3.9	533	17	BH872432	BH872432 hpo5f02.b
C 44	43	3.9	554	10	AM648111	AM648111 EST326565
C 45	42.6	3.8	440	12	BE922932	BE922932 EST426701

ALIGNMENTS

RESULT 1
LOCUS CNS06N13/c
DEFINITION T7 end of clone AU0A008F10 of library AU0A from strain CBS 3082
of Saccharomyces kluyveri, genomic survey sequence.
ACCESSION AL406189.1 GI:12169887
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
2 (bases 1 to 1011)
Neuvellet, C., Bon, E., Lepingle, A., Wincker, P., Artiguenave, F.,
Galliard, C., and Casaregola, S.
Genomic exploration of the hemiascomycetous yeasts: 9.
Saccharomyces kluyveri

JOURNAL MEDLINE 20584719
PUBMED 11152884
REFERENCE 3 (bases 1 to 1011)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage, 2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail : seque@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT This GSS is part of a random genomic sequencing program of thirteen yeast species: *Saccharomyces bayanus* var. *uvarum*, *Saccharomyces exiguus*, *Saccharomyces servazzii*, *Zygosaccharomyces rouxii*, *Saccharomyces kluyveri*, *Kluyveromyces thermotolerans*, *Kluyveromyces lactis* var. *lactis*, *Kluyveromyces marxianus* var. *marxianus*, *Pichia angusta*, *Debaryomyces hansenii* var. *hansenii*, *Pichia sorbitophila*, *Candida tropicalis* and *Yarrowia lipolytica*. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.

FEATURES
source
1..1011
/organism="Saccharomyces kluyveri"
/strain="CBS 3082"
/db_xref="taxon:4934"
/clone="NAUDAA008P10"
/clone_11b="NAUDAA"
/note="end : 17"
complement(185..>1007)
/note="similar to *Saccharomyces cerevisiae* ORF YNLJ34c [similarity to C.carbonum tox d gene]
1 putative frameshift(s)"
/evidence=not_experimental
complement(168..>1007)
/note="similar to *Saccharomyces cerevisiae* ORF YCR102c [similarity to C.carbonum tox d gene]
1 putative frameshift(s)
similar to *Saccharomyces cerevisiae* ORF YLR460c [similarity to C.carbonum tox d protein]"
/evidence=not_experimental

misc_feature
complement(185..>1007)
/note="similar to *Saccharomyces cerevisiae* ORF YNLJ34c [similarity to C.carbonum tox d gene]
1 putative frameshift(s)"
/evidence=not_experimental

misc_feature
complement(168..>1007)
/note="similar to *Saccharomyces cerevisiae* ORF YCR102c [similarity to C.carbonum tox d gene]
1 putative frameshift(s)
similar to *Saccharomyces cerevisiae* ORF YLR460c [similarity to C.carbonum tox d protein]"
/evidence=not_experimental

BASE COUNT 271 a 222 c 215 g 295 t 8 others

ORIGIN

Query Match 26.5%; Score 294.4; DB 17; Length 1011;
Best Local Similarity 61.3%; Pred. No. 8e-67;
Matches 504; Conservative 2; Mismatches 314; Indels 2; Gaps 2;

292 CTTAAAAAAGTGCATTTGCTGAAATGACGAGGCTTATTCACCTTTGTTTACAAGAGT 351
1004 CTTGAGATGATGCTTGCTTGCTGAAATACAGCCGATTTGTCATTTCTTGCTTATTAAGCA 945
352 AACTTAACCTCACTCAACTGCTGATGAATTTCTGAAGCCCTGTGAAGAACTTGCATCT 411
944 GAGTTAACACCTAGTACMAAGRAT-ACATPACAGAAAGGTGTGCTCAAAAACCTTGAATCG 886
412 GCTGCATATTGCTCCAGTTTCGTTGCAACTGTGTGTAGTTGTTGTGCATCATTGGGC 471
885 GACTB-TCGTCACCTGTGCTTCGTCACAGGCTGTGTGTGCTTCAAGTCATCACCTGGGC 827
472 TCAAAAATAGGAATGGCAACCATCTACCCCGCAACATCAATTCATTTATGATTTGGGGT 531
826 AGCAAGTTGGAATGGGAACCAAGCCTCAACATGATTTCCCTCTCTTAATTTGGGGT 767
532 GGTGTACAGAGTGGGTCAACAATATTCAGATTTGCAACAAATATCAATGCTTATCTACT 591
766 GGTGTACAGAGTGGGTCAATTTTGAATTCAGTTGGTGAAGAAATGATACGTTTACCA 707
592 AAGATTTGAATCTGTGCTTTCTTAAAAAGATGAAAAGCTTTAAAGTCTTATGCTGAT 651
706 AAGATTTGGGGTGTGCTTCCAGAAACACAAATCATCTTGAAGCAATACGCTGAT 647
652 GATGCTTTGCTATCAGATGAGCAGCGCTTTATGAGCAGATCAATGCAATATTCACCAAC 711
646 GAGATCTTCATATTCACATGATGAGAGCGTTATTGAACAAATTAAGAGCCAGATGCCAAAT 587

QY	712	CTGCAACATGTTATTGACGCTGTGGGAAGGAATGTAATCCCGAGGCGCTATAAAGTC	772
QY <td>586</td> <td>CTACACACCTGTATCGATGCTGTTTCTTAAGCTGCAATCTATCAACAGGCTTACAGATGT</td> <td>527</td>	586	CTACACACCTGTATCGATGCTGTTTCTTAAGCTGCAATCTATCAACAGGCTTACAGATGT	527
QY <td>772</td> <td>ACACGATAGTCTACCTCCGACATTTATTAAGATGGTCCATGACCATTTGAAGCAAT</td> <td>831</td>	772	ACACGATAGTCTACCTCCGACATTTATTAAGATGGTCCATGACCATTTGAAGCAAT	831
QY <td>832</td> <td>CTGGAAGAAATCAGAAAAGATATATGTTAAATTTATTTACTTTGTGTATCGCATCT</td> <td>891</td>	832	CTGGAAGAAATCAGAAAAGATATATGTTAAATTTATTTACTTTGTGTATCGCATCT	891
QY <td>466</td> <td>GAAAAAGAGAAAGAAATCAAAAGTTAAGATGAGACGCGCTTGTAATCTTGGCTGCC</td> <td>407</td>	466	GAAAAAGAGAAAGAAATCAAAAGTTAAGATGAGACGCGCTTGTAATCTTGGCTGCC	407
QY <td>892</td> <td>GGTCAGAAATTTCTATTGGGTGCAACAAGATTTCTCTGTGTGTCAGAAATATCATGAAGCC</td> <td>951</td>	892	GGTCAGAAATTTCTATTGGGTGCAACAAGATTTCTCTGTGTGTCAGAAATATCATGAAGCC	951
QY <td>406</td> <td>GGCTACGAGTGCCTATTCGCGCCGCCACATCTATCCCTCTAACCGAATACAGAGAGCG</td> <td>347</td>	406	GGCTACGAGTGCCTATTCGCGCCGCCACATCTATCCCTCTAACCGAATACAGAGAGCG	347
QY <td>952</td> <td>ACAGTTAATTCGTTAAGTTATTAATCCACCTTAACAACGGTGTATTCATCATATG</td> <td>1011</td>	952	ACAGTTAATTCGTTAAGTTATTAATCCACCTTAACAACGGTGTATTCATCATATG	1011
QY <td>346</td> <td>ACTGAAGATTCGTGGAATTTGTCAACCCAAAGATCAACAACGGTGAACCTTACCATTTCT</td> <td>287</td>	346	ACTGAAGATTCGTGGAATTTGTCAACCCAAAGATCAACAACGGTGAACCTTACCATTTCT	287
QY <td>1012</td> <td>AATTTTAAGTTTTCAGCAACGCTTGATGATGTCCCACTCTCATGGAAGTATATAA</td> <td>1071</td>	1012	AATTTTAAGTTTTCAGCAACGCTTGATGATGTCCCACTCTCATGGAAGTATATAA	1071
QY <td>286</td> <td>CTAATCCGCTGTTTGAACCGGGTTGGAAGATGTTCTGCACTTACTGCTGCATCA</td> <td>227</td>	286	CTAATCCGCTGTTTGAACCGGGTTGGAAGATGTTCTGCACTTACTGCTGCATCA	227
QY <td>1072</td> <td>GAAGTAAAAACAAAAATGTTAAGTATGTTGCCAGGTATATA</td> <td>1113</td>	1072	GAAGTAAAAACAAAAATGTTAAGTATGTTGCCAGGTATATA	1113
QY <td>226</td> <td>GCAGTTAAGATTCGGGTGTCATTTGTTCCAGCTTTGAAA</td> <td>185</td>	226	GCAGTTAAGATTCGGGTGTCATTTGTTCCAGCTTTGAAA	185
Db			
RESULT 2			
CNS06LDM/c	896 bp	DNA	linear
LOCUS			
DEFINITION	T3 end of clone AT0A012G10 of library AT0A from strain CBS 4311		
ACCESSION	AL404058		
VERSION	AL404058.1		
KEYWORDS	GI:12165095		
SOURCE	GSS.		
ORGANISM	Saccharomyces servazzii.		
REFERENCE	Saccharomyces servazzii.		
AUTHORS	Saccharomycetales; Saccharomycetaceae; Saccharomycetes.		
	1 (bases 1 to 896)		
	Souciek,J.L., Aigle,M., Artiguenave,F., Blandin,G.,		
	Boitard-Pukhara,M., Bon,E., Brodter,P., Caesaregola,S.,		
	De-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,		
	Malpertuy,A., Neuvéglise,C., Ozier-Kalogeropoulos,O., Potier,S.,		
	Saurin,M., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,		
	Wincker,P. and Weissenbach,J.		
	Genomic exploration of the hemiascomycetous yeasts: 1. A set of		
	yeast species for molecular evolution studies		
	FEBS Lett. 487 (1), 3-12 (2000)		
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
COMMENT			

angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila, Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.

FEATURES

Location/Qualifiers

1..896

/organism="Saccharomyces servazzii"

/strain="CBS 4311"

/db_xref="taxon:27293"

/clone="AT0A012G10"

/clone_11b="AT0A"

/note="end : 13"

misc_feature

/note="similar to Saccharomyces cerevisiae ORF YNL134c [

complement(<282..>892)

/note="similar to C.carbonum toxd gene]

/evidence=not_experimental

/evidence=not_experimental

/evidence=not_experimental

/evidence=not_experimental

/evidence=not_experimental

/evidence=not_experimental

/evidence=not_experimental

/evidence=not_experimental

/evidence=not_experimental

/evidence=not_experimental

/evidence=not_experimental

/evidence=not_experimental

/evidence=not_experimental

/evidence=not_experimental

/evidence=not_experimental

/evidence=not_experimental

/evidence=not_experimental

/evidence=not_experimental

/evidence=not_experimental

/evidence=not_experimental

/evidence=not_experimental

/evidence=not_experimental

/evidence=not_experimental

/evidence=not_experimental

/evidence=not_experimental

/evidence=not_experimental

/evidence=not_experimental

/evidence=not_experimental

/evidence=not_experimental

/evidence=not_experimental

/evidence=not_experimental

/evidence=not_experimental

/evidence=not_experimental

/evidence=not_experimental

/evidence=not_experimental

/evidence=not_experimental

/evidence=not_experimental

/evidence=not_experimental

/evidence=not_experimental

/evidence=not_experimental

/evidence=not_experimental

/evidence=not_experimental

/evidence=not_experimental

/evidence=not_experimental

/evidence=not_experimental

/evidence=not_experimental

/evidence=not_experimental

/evidence=not_experimental

/evidence=not_experimental

/evidence=not_experimental

/evidence=not_experimental

/evidence=not_experimental

/evidence=not_experimental

/evidence=not_experimental

/evidence=not_experimental

/evidence=not_experimental

/evidence=not_experimental

/evidence=not_experimental

DEFINITION T7 end of clone AT0A012G10 of library AT0A from strain CBS 4311 of Saccharomyces servazzii, genomic survey sequence.

ACCESSION

AL0404057.1 GI:12165093

VERSION

1

KEYWORDS

GSS.

SOURCE

Saccharomyces servazzii.

ORGANISM

Saccharomyces servazzii

REFERENCE

1 (bases 1 to 891)

Soulier, J. L., Aigle, M., Artiguenave, F., Blandin, G.,

Boitlot, F., Bon, E., Broclet, P., Casaregola, S.,

de Montigny, V., Dujon, B., Durieux, P., Lepingle, A., Llorente, B.,

Malpertuy, A., Neuvéglise, C., Olier-Kalogeropoulos, O., Potier, S.,

Saurin, W., Tekala, F., Toffano-Nicche, C., Wesolowski-Louvel, M.,

Wincker, P. and Weissenbach, J.

Genomic exploration of the hemiascomycetous yeasts: 1. A set of

yeast species for molecular evolution studies

FEBS Lett. 487 (1), 3-12 (2000)

11152876

2 (bases 1 to 891)

Caesaregola, S., Lepingle, A., Bon, E., Neuvéglise, C., Nguyen, H.,

Artiguenave, F., Wincker, P. and Galliard, C.

Genomic exploration of the hemiascomycetous yeasts: 7.

Saccharomyces servazzii

FEBS Lett. 487 (1), 47-51 (2000)

11152882

3 (bases 1 to 891)

Genoscope.

Direct Submission

Submitted (07-SEP-2000) Genoscope - Centre National de Séquençage,

2 rue Gaston Cremieux, CP 5706, 91057 Evry cedex, FRANCE. (E-mail :

segret@genoscope.cns.fr - Web : www.genoscope.cns.fr)

This GSS is part of a random genomic sequencing program of thirteen

yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces

exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,

Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces

lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia

angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,

Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to

5 kb were prepared and both extremities were sequenced. See

keywords for description of this sequence and for the sequence of

the other extremity of this insert.

Location/Qualifiers

1..891

/organism="Saccharomyces servazzii"

/strain="CBS 4311"

/db_xref="taxon:27293"

/clone="AT0A012G10"

/clone_11b="AT0A"

/note="end : 17"

<140..>877

/note="similar to Saccharomyces cerevisiae ORF YLR460c [

similarity to C.carbonum toxd protein]"

/evidence=not_experimental

/evidence=not_experimental

/evidence=not_experimental

/evidence=not_experimental

/evidence=not_experimental

/evidence=not_experimental

/evidence=not_experimental

/evidence=not_experimental

/evidence=not_experimental

/evidence=not_experimental

/evidence=not_experimental

/evidence=not_experimental

/evidence=not_experimental

/evidence=not_experimental

/evidence=not_experimental

/evidence=not_experimental

BASE COUNT 236 a 231 c 205 g 209 t

ORIGIN

1

ATGTCAGTTCACCACTCAAAAGCGCTCATTTGAAGTGACAAAGCTGTGTTAA 60

Query Match

Best Local Similarity 54.7%; Pred. No. 5.7e-37; Indels 6; Gaps 2;

Matches 404; Conservative 0; Mismatches 328

1

ATGTCAGTTCACCACTCAAAAGCGCTCATTTGAAGTGACAAAGCTGTGTTAA 60

Query Match

Best Local Similarity 54.7%; Pred. No. 5.7e-37; Indels 6; Gaps 2;

Matches 404; Conservative 0; Mismatches 328

1

ATGTCAGTTCACCACTCAAAAGCGCTCATTTGAAGTGACAAAGCTGTGTTAA 60

Query Match

Best Local Similarity 54.7%; Pred. No. 5.7e-37; Indels 6; Gaps 2;

Matches 404; Conservative 0; Mismatches 328

1

ATGTCAGTTCACCACTCAAAAGCGCTCATTTGAAGTGACAAAGCTGTGTTAA 60

Query Match

Best Local Similarity 54.7%; Pred. No. 5.7e-37; Indels 6; Gaps 2;

Matches 404; Conservative 0; Mismatches 328

1

ATGTCAGTTCACCACTCAAAAGCGCTCATTTGAAGTGACAAAGCTGTGTTAA 60

Query Match

Best Local Similarity 54.7%; Pred. No. 5.7e-37; Indels 6; Gaps 2;

Matches 404; Conservative 0; Mismatches 328

1

ATGTCAGTTCACCACTCAAAAGCGCTCATTTGAAGTGACAAAGCTGTGTTAA 60

Query Match

Best Local Similarity 54.7%; Pred. No. 5.7e-37; Indels 6; Gaps 2;

Matches 404; Conservative 0; Mismatches 328

1

ATGTCAGTTCACCACTCAAAAGCGCTCATTTGAAGTGACAAAGCTGTGTTAA 60

Query Match

Best Local Similarity 54.7%; Pred. No. 5.7e-37; Indels 6; Gaps 2;

Matches 404; Conservative 0; Mismatches 328

1

ATGTCAGTTCACCACTCAAAAGCGCTCATTTGAAGTGACAAAGCTGTGTTAA 60

Query Match

Best Local Similarity 54.7%; Pred. No. 5.7e-37; Indels 6; Gaps 2;

Matches 404; Conservative 0; Mismatches 328

1

ATGTCAGTTCACCACTCAAAAGCGCTCATTTGAAGTGACAAAGCTGTGTTAA 60

Query Match

Best Local Similarity 54.7%; Pred. No. 5.7e-37; Indels 6; Gaps 2;

Matches 404; Conservative 0; Mismatches 328

1

ATGTCAGTTCACCACTCAAAAGCGCTCATTTGAAGTGACAAAGCTGTGTTAA 60

Query Match

Best Local Similarity 54.7%; Pred. No. 5.7e-37; Indels 6; Gaps 2;

Matches 404; Conservative 0; Mismatches 328

1

ATGTCAGTTCACCACTCAAAAGCGCTCATTTGAAGTGACAAAGCTGTGTTAA 60

Query Match

Best Local Similarity 54.7%; Pred. No. 5.7e-37; Indels 6; Gaps 2;

Matches 404; Conservative 0; Mismatches 328

LOCUS	DEFINITION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS
Db	140	ATGTCCTTACCACTCCATGCAAGCCACCGTCAATCGGTAAGCCGTCATCAAA	199				
Oy	61	AACAGATGCTCAATTCACGAATTTAAAGAGGGTACAGCTTGGTGAAGTTGAGGCTGT	120				
Db	200	AC---CGTCCCATGCTCTCAACGTGACGACGAAATACGTCCTTGATCAAGACAAAGCCGT	256				
Oy	121	GCTGGTAAACCAACGATTTGGAGAGCATTTGCTTAAAGATTGGTCCAGAAAGTTCAAT	180				
Db	257	GCCGCTAAACCAACGATTTGGAGAGCATTTGCTTAAAGATTGGTCCAGAAAGTTCAAT	316				
Oy	181	CTAGAGATGTCATATGCTGTGTACAGTTGTCAAACTTGGACCAAAATGCTAGTACTGAC---	237				
Db	317	ATCGGTGTAGATGCCGCCGGTGAATTCGTCAAGTTAGTTTCCAAAGTGCACAAACAA	376				
Oy	238	TTGAAGTTGAGATACCGGTTTGGTTTGTTCACGGTGTCTCCAAACAGATCTTAA	297				
Db	377	TTCCACGTAGAGATATATGTCTACGGGTTGCTTCAATGAGTCTTCAAGAGACGTCAGAT	436				
Oy	298	AATGTGATTTTGTCTGATATATGCCAGGGTTATCCACTTGTGTTTACAAAGTAACTTA	357				
Db	437	AATGGGGCTTTGACGAATACGTTGCTCTAGATTTCATATCACCTTCCCTGTCCTCCAC	496				
Oy	358	ACTCACTCAACTGCGATGATGAATTTCTGAAAGGCCCTGTGAAGAACTTGAAATCTGCTCA	417				
Db	497	GCTCTTTCGGCAGAGATTCATTCATCAAGGGCCCGGTGACCTCTTGAAAGCCGGTCC	556				
Oy	418	TCATTTGCAATTTGCTGTGACCAACTGCTGTGTATTGTTGTGTCATCACTTGGGCTCAAA	477				
Db	557	ACGATTCATGTCCTGCTGTGACGTCAGGGCCGCACTTATTCACCAATGGGTCTCAAC	616				
Oy	478	ATGGAATGACCACTTCTACCCCGCAACTACTATTCATTAATTTGGGGTGTGTCT	537				
Db	617	ATGAGTGTGGGAACCAAGCAGGTCCAAAGAAAGGGAACGTGTTGATCTGGGGGGGTGCT	676				
Oy	538	ACAGCAGTGGGTCACCAACTATTCACAGTTGCAACATATCAATGCTTATTAATTAAGAT	597				
Db	677	ACCGCTCTGTCTCTTGGCACTTTACACTATTTGCAAAAAGTTAATGCTTGGAGAGTCT	736				
Oy	598	GTAACGTGTGCTTTTAAAGAGATGAAAGCTTTTAAAGCTTTATGTGTGATGATGATGTC	657				
Db	737	ATTGTCGTGGGCTCCAGAAACATGAACTAATTTAAAGAAATACGGTGTGATGATGATTA	796				
Oy	658	TTTGAATATCATGATGACAGGGCTTATTTGACGATCAATATGAAATATCAAACTTGCA	717				
Db	797	TTGCACTATTCAGATCCGAGCTTATTTGACCAATCAAGGCCAAATACACGATTTGGTT	856				
Oy	718	CATGTTATTTGACGCTGTG 735					
Db	857	TGCTTATTTAGATTGTGTG 874					
LOCUS	CNS06LTC	1072 bp	DNA	linear	GSS 17-JUN-2001		
DEFINITION	T7 end of clone XAT0A001D07 of library XAT0A from strain CBS 43111						
VERSION	AL404614						
KEYWORDS	AL404614.1 GI:12166302						
SOURCE	GSS.						
ORGANISM	Saccharomyces servazzii.						
REFERENCE	Saccharomyces servazzii.						
AUTHORS	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces.						
	1 (bases 1 to 1072)						
	Soucieur, J. L., Aisle, M., Artiguenave, F., Blandin, G.,						
	Botolich, F., Kuhn, M., Bon, B., Brodier, P., Casaregola, S.,						
	de-Montigny, V., Dujon, B., Durieux, P., Lepingle, A., Llorente, B.,						
	Malpertuy, A., Neuvéglise, C., Ozier-Kalogeropoulos, O., Potier, S.,						
	Saurin, M., Tekaia, F., Toffano-Nicolas, C., Wesolowski, L.,						
	Wincker, P., and Weissenbach, J.						
	Genomic exploration of the hemiascomycetous yeasts: 1. A set of						
	yeast species for molecular evolution studies						

FEATURES	source
JOURNAL MEDLINE	FEMS Lett. 487 (1), 3-12 (2000)
PUBMED	20584711
REFERENCE	11152876
AUTHORS	2 (bases 1 to 1072)
TITLE	Caaregola, S., Lepingle, A., Bon, E., Neuveglise, C., Nguyen, H., Ariguenave, F., Wincker, P., and Gallardin, C. Genomic exploration of the hemiascomycetous yeasts: 7. <i>Saccharomyces servazii</i>
JOURNAL MEDLINE	FEMS Lett. 487 (1), 47-51 (2000)
PUBMED	20584717
REFERENCE	11152882
AUTHORS	3 (bases 1 to 1072)
TITLE	Genoscope. Direct Submission
JOURNAL	Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage, 2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr) This GSS is part of a random genomic sequencing program of thirteen yeast species: <i>Saccharomyces bayanus</i> var. <i>uvaurum</i> , <i>Saccharomyces exiguus</i> , <i>Saccharomyces servazii</i> , <i>Zygosaccharomyces rouxii</i> , <i>Saccharomyces kluyveri</i> , <i>Kluyveromyces marxianus</i> var. <i>marxianus</i> , <i>Pichia angusta</i> , <i>Debaryomyces hansenii</i> var. <i>hansenii</i> , <i>Pichia sorbitophila</i> , <i>Candida tropicalis</i> and <i>Yarrowia lipolytica</i> . Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert. Location/Qualifiers
misc_feature	1..1072 /organism="Saccharomyces servazii" /strain="CBS 4311" /db_xref="taxon:27293" /clone="XAT0A001D07" /clone_11b="XAT0A" /note="end : T7" <336..>1039 /note="similar to <i>Saccharomyces cerevisiae</i> ORF YLR460c [similarity to C.carbonum toxo protein] 2 putative frameshift(s) similar to <i>Saccharomyces cerevisiae</i> ORF YNL134c [similarity to C.carbonum toxo gene]" /evidence=not_experimental <348..>1039 /note="similar to <i>Saccharomyces cerevisiae</i> ORF YCR102c [similarity to C.carbonum toxo gene] 2 putative frameshift(s)" /evidence=not_experimental
misc_feature	
BASE COUNT	350 a 178 c 210 g 331 t 3 others
ORIGIN	
Query Match	13.9%; Score 154.6; DB 17; Length 1072;
Best Local Similarity	56.0%; Pred. No. 6.2e-30;
Matches 394;	Conservative 1; Mismatches 270; Indels 38; Gaps 4;
QY	8 TTCCAACCTGCAGAAAAGCCGCTCATCTTAAGAGTGACAAAGCTGTTTAAACAGATG 67
DB	337 TTCCAAGACATGAAAGCCGCTCTTTTAAAGGACAGATTGCAGAAAAGTTCAAGATAGA 396
QY	68 TCTCAGTTTCAGATTAAAGAGGGTACACCTTGAGAGTTGAGGCTGTGGCTGTA 127
DB	397 TTAAAGTTCCCTTGCTGATGATGATGGTTCCTTTAATTCAGATTTCAGAGTGGACACA 456
QY	128 ACCCACTGATGGAGACATATTGCTTAAAGATTGTCAGAAAGTTTCAATTCTAGAT 187
DB	457 ATCCAAATGATTTGAAACATGTTGAATTTGGAATGGGTCCAAAAGGTCACATCTTAGTT 516
QY	188 GTGACATTGCTGTGATCAGTTGTCAAATTGGACCA--AATGCTAGTACGACTTGAAG 244
DB	517 GTGATGTTGCGGTTGTGTGTGTGCTTATTTGGGTCAGGAAATTCACGACATGAATATAAG 576
QY	245 TTGAGAGTACCAGTTTGGGTTTGGTTTTCACGGATGTTCCCAACAGATCCTAAAAATGTTG 304
DB	577 TTGGTGAATATGTCGTGACATTGTTCTATAGGGGGGTATTCCTGTAACCTGGCAATGGG 636

Qy 305 CATTGCTGATATGCGACAGGTTTATCCACTTTGTTT----- 343
 Db 637 CATTGCGACAAATTTGTGCTGATGATGCAAAATATCACTTAATTTACCTGAAACATGA 696
 Qy 344 -----ACAAGACTTAATTAATCTCACTCACTGCTGATGAAATTTCTGAAGGC 391
 Db 697 TGCTTCATAGACACAGGCTTGGAAACAGGTCAAGATTTATTCAGATGATGAAAGGC 756
 Qy 332 CTGTGAAGAACTTCCAAATGCTGCATCTTGGCACTTTGTTGACAACTGCTGCTTGA 451
 Db 757 CAGTGAATTTCTATTGAACTGCTGCTAATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 816
 Qy 452 GTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 511
 Db 817 TTTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 876
 Qy 512 ATCCATTTATTTGTTT-GGGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 570
 Db 877 CACCACTTTTATTTGGGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 936
 Qy 571 AAACATATCAATGCTTATTAATGATTTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 629
 Db 937 AAAAATTAATGATGCTTATTAATGATTTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 996
 Qy 630 TTTAAAGTCTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 672
 Db 997 GCTAATCATTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1039

RESULT 5
 AZ931755/c 488 bp DNA linear GSS 01-APR-2001
 LOCUS AZ931755

DEFINITION 474.dh89f04.s1.Saccharomyces unisporus NRRL Y-1556 Saccharomyces
 unisporus genomic clone 474.dh89f04.s1, DNA sequence.

ACCESSION AZ931755
 VERSION AZ931755.1 GI:13502666

KEYWORDS GSS.
 SOURCE Saccharomyces unisporus.

ORGANISM Saccharomyces unisporus.
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

REFERENCE 1 (bases 1 to 488)
 Saccharomycetales; Saccharomycetaceae; Saccharomyces.

AUTHORS Cliften,P.F., Hillier,L.W., Fulton,L., Graves,T., Miner,T., Gish
 W.R., Waterston,R.H. and Johnston,M.
 Surveying Saccharomyces genomes to identify functional elements by
 comparative DNA sequence analysis

TITLE Unpublished (2001)
 JOURNAL Contact: Johnston M

DEPARTMENT OF GENETICS
 Washington University Medical School
 Box 8232, 4566 Scott Ave., St. Louis, MO 63110, USA
 Tel: 314 362 2735
 Fax: 314 362 7855

EMAIL: mj@genetics.wustl.edu
 CLASS: random plasmid subclone.
 FEATURES
 source

1.488
 location/Qualifiers

/organism="Saccharomyces unisporus"
 /strain="NRRL Y-1556 (CBS 398)"
 /db_xref="taxon:27294"
 /clone="474.dh89f04.s1"
 /clone_lib="Saccharomyces unisporus NRRL Y-1556"
 /note="Random genomic sequence"

BASE COUNT 154 a 97 c 79 g 158 t
 ORIGIN

Query Match 13.2% Score 146.6; DB 17; Length 488;
 Best Local Similarity 58.2%; Pred. No. 6.7e-28;
 Matches 219; Conservative 0; Mismatches 194; Indels 6; Gaps 1;
 256 GGTTCGATTTTCTCAAGGCTTCCCAACAGATCCCTAAATGCTGCTGCTGCTGCTGCTGCTG 315

Db 482 GTTTCGCTATTTGTTAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 423
 Qy 316 TATGCGAGGTTTATCCACCTTTGTTTAAAG-----AGTAATTAATCAGCTCACT 369
 Db 422 TACGGAAGAGGATTCGAATATCTGCTTCAAGTCCCAAAATTAATTAATTAATCTTAC 363
 Qy 370 GCTGATGAATTTTCTGAAGCCCTGCTGGAAGAACTTCAATCTGCTGCTGCTGCTGCTGCTG 429
 Db 362 AAGATTAATATCGAGGACAGTAAATTTGACACTTGGGAAGCTTGCTGCTGCTGCTGCTGCTG 303
 Qy 430 TCGTTGACAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 489
 Db 302 GTATTTGATATCCGCTTTAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 243
 Qy 430 CCATTCACCGGCAATATCATTCATTTATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTG 549
 Db 242 CCAAGAAAACCAACATGATCATCATTTAATTTGGGGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 183
 Qy 550 CAACAATTAATCAAGTTCGCAACATATCATTCATTTATTAAGATTTGATTTGATTTGATTTGATTTG 609
 Db 182 CAATATGCTATTTCAATTTGCTCAAGCAATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTT 123
 Qy 610 TCTAAAGCAATGAAGCTTTTAAAGTCTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 669
 Db 122 TCAAGAAAACATGAAGAAATGTTAAATCTTTGCTGCAATGAAGTCTTCAATATCAT 63
 Qy 670 GATGAGGCTTATTTGACAGATCAATCGAAGTTCCAAACCTGCAACATGTTATGA 728
 Db 62 GACAAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4

RESULT 6
 CDS06LW5/c 1118 bp DNA linear GSS 17-JUN-2001

LOCUS CDS06LW5
 DEFINITION T3 end of clone XAT0A002A04 of library XAT0A from strain CBS 4311
 of Saccharomyces servazii, genomic survey sequence.

ACCESSION AL404715
 VERSION AL404715.1 GI:12166529

KEYWORDS GSS.
 SOURCE Saccharomyces servazii.

ORGANISM Saccharomyces servazii.
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

REFERENCE 1 (bases 1 to 1118)
 Saccharomycetales; Saccharomycetaceae; Saccharomyces.

AUTHORS Soucier,U.L., Aigle,M., Artiguenave,F., Blandin,G.,
 de-Montigny,V., Dujon,B., Durieux,P., Lepingle,A., Lorente,B.,
 Malpertuy,A., Neuvéglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
 Saurin,W., Tekala,F., Toffano-Nicolas,C., Wesolowski-Louvel,M.,
 Wincker,P. and Weissenbach,J.

TITLE Genomic exploration of the hemiascomycetous yeasts: 1. A set of
 yeast species for molecular evolution studies
 JOURNAL FEMS Lett. 487 (1), 3-12 (2000)

MEDLINE 20584711
 PUBMED 1152876

REFERENCE 2 (bases 1 to 1118)
 Artiguenave,F., Wincker,P. and Galliardin,C.
 Genomic exploration of the hemiascomycetous yeasts: 7.

JOURNAL FEMS Lett. 487 (1), 47-51 (2000).
 MEDLINE 20584717

PUBMED 1152882
 JOURNAL 3 (bases 1 to 1118)

TITLE Genoscope.
 JOURNAL Direct Submission

AUTHORS Submitted (07-SEP-2000) Genoscope - Centre National de Séquençage,
 2 rue Gaston Crémieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail:

segre@genoscope.cns.fr - Web: www.genoscope.cns.fr)
 COMMENT This GSS is part of a random genomic sequencing program of thirteen

yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
 exiguus, Saccharomyces servazii, Zygosaccharomyces rouxi,
 Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces

lactis var. *lactis*, *Kluyveromyces marxianus* var. *marxianus*, *Pichia angusta*, *Debaryomyces hansenii* var. *hansenii*, *Pichia sorbitophila*, *Candida tropicalis* and *Yarrowia lipolytica*. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.

FEATURES

source
1. .1118
/organism="Saccharomyces servazii"
/strain="CBS 4311"
/db_xref="taxon:27293"
/clone="XAT0A002A04"
/clone_1id="XAT0A"
/note="end : T3"
misc_feature
/note="similar to *Saccharomyces cerevisiae* ORF YNL134c [similarity to C.carbonum tox gene]
2 putative frameshift(s)"
/evidence=not_experimental
misc_feature
/note="similar to *Saccharomyces cerevisiae* ORF YCR102c [similarity to C.carbonum tox gene]
2 putative frameshift(s)"
/evidence=not_experimental
BASE COUNT 366 a, 217 c 202 g 328 t 5 others
ORIGIN

Query Match

Best Local Similarity 54.1%; Pred. No. 2,4e-25; Length 1118;
Matches 379; Conservative 1; Mismatches 294; Indels 26; Gaps 4;

QY 11 CAACTCACTCAAAAGCCGTCATCTGAGTGAAGTCAAGCTGTTGTAAGACAGATGCT 70
DB 1112 CAACCTCTAAGAAGCCACCGTATCGACAATGTAATCTCTATCAAAAC---TGTC 1056
QY 71 CATTCCAGATTAAGAGAGGTACAGCCCTTGAGAGGTGAGCTGTGCTGTAAC 130
DB 1055 CATTGCCACACGTGAGAGAGATACGCTTTAACAGAGCAATATCATGCTGTAAC 996
QY 131 CAATGATGGAAGCATATATGCTTAATAGATGGTCCAGAGGTCAATCTAGAGATG 190
DB 995 CGACGGGTGAAGACGTAGCAAGTATGCTTGGGCCAATGGTCCGTATGGTGG 936
QY 191 ACATTCCTGTGATACGTTGCAAACTTGACCAATG---CTAGTACTGACTTGAAGT 247
DB 935 ATCCCTCTGTTGAATGTGAATAGCTTCTAAGCTATACAACTTAATTCACCTAG 876
QY 248 GAGATACCGGTTGGTTTGTCAAGGTGCTCCCAAGACAGATCCCAAAAGGTGCAT 307
DB 875 GTGACTGTGTTAATGTTTATTCATGCTCTTCCAGAGAGCTCGACACGGTGCAT 816
QY 308 TTGCTGAATATGACAGGGTTATTCACCTTTGTTTACAGAGTAATTAATCACTCA 367
DB 815 TTGCCAGTACGCTGCTGTATTTCTAAGATCTCATTTATA----- 774
QY 368 CTGCTGATGAATTTCTGAAGCCCTGTGAAGAACTTGAATCTGCTGATCAATGGCAG 427
DB 773 -TGAAATAGGGGTTTTCAGGTGAAGTTTCACTCCCGAAGGTGCTGCTGATCCAT 715
QY 428 TTTCGTTGCAACTGCTGCTGTTAGTTTGTGTCATCATCTGGGCTCAAAATGGAATG 487
DB 714 CGTCTGTGTTGACAGCTTCTGCAACTCTCTTTCT-ACAGCTGCTTTAAAGGAATG 656
QY 488 ACCATCTACCCGCAACATATCATCTCATTTATGATTTGGGTGCTGACAGAGTG 547
DB 655 AACCAAGAGAGGCAATAAAGTATCTGTTTMAATTTGGATGCTGCTGCTTGT 596
QY 548 GTCAACATTAATCAAGTTCAGCAACATATCAATGCTTAATCAATGTTGAATGTTG 607
DB 595 CTCTAGCATTTTCAATTAATTCACAAATTCATATGTTGCAAAATGTTGCTGTTG 536
QY 608 CTCTTAAAGCATGAAGACCTTTTAAGCTTAATGCTGTAAGTCTTTGACATAC 667

DB 535 CCTCGAAGAACGTGAACCTAATCTAAGAAATGGGTCTACCTGAGATTTGATTAT 476
QY 668 ATGATCAGGCGCTTATTTGAGACATCAATCAAGATATCC 707
DB 475 ATGATCCTATATTTGTTGAACAATCAATCAATCAATATATAC 436

RESULT 7
A2931754 434 bp DNA linear GSS 01-APR-2001
LOCUS 474.dhz89f03.s1 Saccharomyces unisporus NRRL Y-1556 Saccharomyces
DEFINITION unisporus genomic clone 474.dhz89f03.s1, DNA sequence.
ACCESSION A2931754
VERSION A2931754.1 GI:13502665
KEYWORDS GSS.
SOURCE Saccharomyces unisporus.
ORGANISM Saccharomyces unisporus.
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

REFERENCE

1 (bases 1 to 434)
Clifton, P.F., Hillier, L.W., Fulton, L., Graves, T., Miner, T., Gish
W.R., Waterston, R.H. and Johnston, M.
Surveying Saccharomyces genomes to identify functional elements by
comparative DNA sequence analysis
Unpublished (2001)

JOURNAL

Contact: Johnston M
Department of Genetics
Washington University Medical School
Box 8237, 4566 Scott Ave., St. Louis, MO 63110, USA
Tel: 314 362 2735
Fax: 314 362 7855

Email: mj@genetics.wustl.edu
Class: random plasmid subclone.
Location/Qualifiers

FEATURES

source

1. .434
/organism="Saccharomyces unisporus"
/strain="NRRL Y-1556 (CBS 338)"
/db_xref="taxon:27294"
/clone="474.dhz89f03.s1"
/clone_1id="Saccharomyces unisporus NRRL Y-1556"
/note="Random genomic sequence"
BASE COUNT 133 a 76 c 88 g 137 t

ORIGIN

Query Match

Best Local Similarity 59.7%; Pred. No. 3.2e-24;
Matches 244; Conservative 0; Mismatches 159; Indels 6; Gaps 1;

QY 256 GGTTCGGTTTGTTCACGCTGCTCCCAAGATCCCTAAATGGTGCATTTGCTGA 315
DB 26 GGTGCTGCTGTGATTAATGTTGATGTTCAATTCACGATTAATGCTTAATCTGA 85
QY 316 TATGCAGGGTTTATTCACCTTTGTTTACAA-----GAGTAATCTAATCACTCACT 369
DB 86 TATGTTAAGTGAACCAACATATTTTCAATTTCCCAAGACATTAATCTTTACTAA 145
QY 370 GCTGATGAATTTCTGAAGCCCTGTGAAGAACTTGAATGCTGTCATCATTTGCCAGT 429
DB 146 AAGACTGACGTGGCAACAGCAAGTTGACACATGGAGAGCTTGGCCAGTTCCCAATG 205
QY 430 TCGTTGACAACTGCTGCTGTTAGTTTGTCTATCACTTGGGCTCAAAATGGAATGGCAG 489
DB 206 GTGTCTTACACGCTTTAGCTGTACCTTTTACCAATGAAATTAATTAAGATGGCA 265
QY 490 CCATTAACCCGCAACATATCTATCATCTATTTGATTTGGGTGCTGACAGAGGTG 549
DB 266 CCGAAGAGCACATATTAACATCAATCTTAATTTGGGGTGGTCCACTGCTGCTGCT 325
QY 550 CAACAATTAATCAAGTTCGCAACATATCAATGCTTAATCAATTAATGTTAACTGTTGCT 609
DB 326 CATATGATTAATCAATTTGTTAAACAGTCAATGGTTTAAAGATTAATGTTGTTGCC 385
QY 610 TCTAAAGACATGAAGACCTTTTAAAGCTTAATGCTGATGATGATCTC 658

Db 386 TCTAAGAAACATGAAAGTTGTAAATCTTAATGCTGCTGAGTAATCT 434

RESULT 8

AO501781/c 454 bp DNA linear GSS 29-APR-1999

LOCUS V14B11 mtN-3xHA/lacZ Insertion Library Saccharomyces cerevisiae

DEFINITION genomic 5', DNA sequence.

ACCESSION AO501781

VERSION AO501781.1 GI:4707431

KEYWORDS GSS.

SOURCE baker's yeast.

ORGANISM Saccharomyces cerevisiae

REFERENCE Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

Saccharomycetales; Saccharomycetaceae; Saccharomyces.

AUTHORS 1 (bases 1 to 454)

REFERENCE Rosse-Macdonald, P., Roemer, T., Coelho, P.S.R., Agarwal, S., Kumar, A.,

deBastges, S.A., Cheung, K.-H., Sheehan, A., Symoniatidis, D., Jansen, R.,

Umansky, L., Heideman, M., Nelson, K., Iwasaki, H., Kanada, D., Lugo, R.,

Hager, K., Miller, P., Roeder, G.S. and Snyder, M.

Large-Scale Analysis of the Yeast Genome by Transposon Tagging and

Gene Disruption

Unpublished (1999)

JOURNAL Contact: Kumar A

COMMENT Michael Snyder, Dept. of Mol. Cell. and Dev. Biology

Yale University

P.O. Box 208103, New Haven, CT 06520-8103, USA

Tel: 203 432 9949

Fax: 203 432 6161

Email: anuj.kumar@yale.edu

te of mtN-3xHA/lacZ insertion.

Seq primer: GGCTTCTTCTTGGAAAGTAC

Class: transposon-tagged.

Location/Qualifiers

1. 454

/organism="Saccharomyces cerevisiae"

/db_xref="taxon:4932"

/clone_1lb="mtN-3xHA/lacZ Insertion Library"

/lab_host="E. coli"

/note="Vector: pHS6-Sal; A yeast genomic DNA library

(lacking mitochondrial DNA) was prepared in pHS6-Sal;

genomic DNA was size-fractionated (DNA of roughly 2-3 kb

in length) prior to cloning. This library was

subsequently mutagenized with a mtN-3xHA/lacZ

mini-transposon containing lacZ, URA3, and tet resistance."

BASE COUNT 122 a 106 c 102 g 124 t

IN

Query Match 11.4%; Score 126.6; DB 17; Length 454;

Best Local Similarity 64.7%; Pred. No. 1.2e-22;

Matches 205; Conservative 0; Mismatches 109; Indels 3; Gaps 1;

Db 4 TCAGTTCACCACTCAAAAGCCGTCATCATGAGTGACAAAGCTGTGTTAAACA 63

Qy 401 TCGATTCCAGAAACCATGAAAGCGTTCATGTAAGGCAAGGCTTATCAACAG 342

Db 64 GATGCTCAGTTCAGAAATTAAGAGGCTACAGCTTGTGAAGCTGAGCTGTGCT 123

Qy 341 GCAATTCACATTCCTGAATTAAGAGAGATTTGTTCTAATTAAGCTGTCCCGTGGC 282

Db 124 GGTAAACCACTGATTTGAAGATATGCTTAATTAAGATGTCGCAAGAGTTCATTCGA 183

Qy 281 GGTAAACCTTACCGATTTGAAGATATGCTTAATTAAGATGTCGCAAGAGTTCATTCGA 222

Db 184 GGAATGACATTCGCTGTGACAGTGTCAACTTGACCAATGCTAGTACTG---ACTTG 240

Qy 221 GGCTGTGATGACAGCGGCAAAATCGTAAAGTTGGGCGCAAGTGTGATCTCAGCGCTT 162

Db 241 AAGGTTGAGATACCGGCTTTGGTTTGTTCACGGTCTTCCCAACAGATCTTAAAT 300

Qy 161 GCCTATGATGATTAATTAATGAGGTTATTAACGAGTGTTCAGTGAGTGTCCCTCAAC 102

Qy 301 GGTGATTTGCTGATTA 317

Db 101 GGTGCTTTGCTGATTA 85

RESULT 9

AO873738/c

LOCUS V74B3 mtN-3xHA/lacZ Insertion Library, strain AB972 Saccharomyces

DEFINITION genomic 5', DNA sequence.

ACCESSION AO873738

VERSION AO873738.1 GI:6285982

KEYWORDS GSS.

SOURCE baker's yeast.

ORGANISM Saccharomyces cerevisiae

REFERENCE Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

Saccharomycetales; Saccharomycetaceae; Saccharomyces.

AUTHORS 1 (bases 1 to 546)

REFERENCE Rosse-Macdonald, P., Roemer, T., Coelho, P.S.R., Agarwal, S., Kumar, A.,

deBastges, S.A., Cheung, K.-H., Sheehan, A., Symoniatidis, D., Jansen, R.,

Umansky, L., Heideman, M., Nelson, K., Iwasaki, H., Kanada, D., Lugo, R.,

Hager, K., Miller, P., Roeder, G.S. and Snyder, M.

Large-Scale Analysis of the Yeast Genome by Transposon Tagging and

Gene Disruption

Unpublished (1999)

JOURNAL Contact: Kumar A

COMMENT Michael Snyder, Dept. of Mol. Cell. and Dev. Biology

Yale University

P.O. Box 208103, New Haven, CT 06520-8103, USA

Tel: 203 432 9949

Fax: 203 432 6161

Email: anuj.kumar@yale.edu

te of mtN-3xHA/lacZ insertion.

Seq primer: GGCTTCTTCTTGGAAAGTAC

Class: transposon-tagged.

Location/Qualifiers

1. 546

/organism="Saccharomyces cerevisiae"

/strain="AB972 - trp1 r(10) (S288C background)"

/db_xref="taxon:4932"

/clone_1lb="mtN-3xHA/lacZ Insertion Library, strain AB972"

/lab_host="E. coli"

/note="Vector: pHS6-Sal; A yeast genomic DNA library

prepared in pHS6-Sal; genomic DNA was size-fractionated

(DNA of roughly 2-3 kb in length) prior to cloning. This

library was subsequently mutagenized with a mtN-3xHA/lacZ

mini-transposon containing lacZ, URA3, and tet resistance."

BASE COUNT 157 a 121 c 114 g 150 t 4 others

ORIGIN

Query Match 11.3%; Score 125.6; DB 17; Length 546;

Best Local Similarity 64.4%; Pred. No. 2.4e-22;

Matches 204; Conservative 0; Mismatches 110; Indels 3; Gaps 1;

Db 4 TCAGTTCACCACTCAAAAGCCGTCATCATGAGTGACAAAGCTGTGTTAAACA 63

Qy 416 TCGATTCCAGAAACCATGAAAGCGTTCATGTAAGGCAAGGCTGTGATCMAAG 357

Db 64 GATGCTCAGTTCAGAAATTAAGAGGCTACAGCTTGTGAAGGTTGAGGCTGTGCT 123

Qy 356 GCAATTCACATTCCTGAATTAAGAGAGATTTGTTCTAATTAAGACGTGCGCTTGC 297

Db 124 GGTAAACCACTGATTTGAAGATATGCTTAATTAAGATGTCGCAAGAGTTCATTCGA 183

Qy 296 GGTAAACCTTACCGATTTGAAGATATGCTTAATTAAGATGTCGCAAGAGTTCATTCGA 237

Db 184 GGAATGACATTCGCTGTGACAGTGTCAACTTGACCAATGCTAGTACTG---ACTTG 240

Qy 236 GGCTGTGATGACAGCGGCAAAATCGTAAAGTTGGGCGCAAGTGTGATCTCAGCGCTT 177

Db 241 AAGGTTGAGATACCGGCTTTGGTTTGTTCACGGTCTTCCCAACAGATCTTAAAT 300

Db 176 GCCATTGGTATTACATTATAGGGGTTATTCAGGCTCTCAGTAGGTTCCCTCAAC 117

QY 301 GGTCATTTCCTGTAATA 317
 |||||
 Db 116 GGTGCTTTCCTGAGTA 100

RESULT 10
 AO873797/c 537 bp DNA linear GSS 08-NOV-1999
 LOCUS V7591 mtn-3xHA/lacZ Insertion library, strain AB972 Saccharomyces
 DEFINITION cerevisiae genomic 5', DNA sequence.
 ACCESSION AO873797
 VERSION AO873797.1 GI:6286041
 KEYWORDS GSS.
 SOURCE baker's yeast.
 ORGANISM Saccharomyces cerevisiae
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 RENCE 1 (bases 1 to 537)
 THORS Ros-Macdonald, P., Roemer, T., Coelho, P. S. R., Agarwal, S., Kumar, A.,
 deStranges, S. A., Cheung, K.-H., Sheehan, A., Symoniatidis, D., Jansen, R.,
 Umansky, L., Heidtman, M., Nelson, K., Iwasaki, H., Kanada, D., Lugo, R.,
 Hager, K., Miller, P., Roeder, G. S. and Snyder, M., Kanada, D., Lugo, R.,
 Large-Scale Analysis of the Yeast Genome by Transposon Tagging and
 Gene Disruption
 Unpublished (1999)
 TITLE Contact: Kumar A
 JOURNAL Michael Snyder, Dept. of Mol. Cell. and Dev. Biology
 COMMENT Yale University
 P.O. Box 208103, New Haven, CT 06520-8103, USA
 Tel: 203 432 9949
 Fax: 203 432 6161
 Email: anuj.kumar@yale.edu
 te of mtn-3xHA/lacZ insertion.
 Seq primer: GGCTTCTTCTTTGGAAAGTAC
 Class: transposon-tagged.
 Location/Qualifiers
 FEATURES
 source 1. 537
 /organism="Saccharomyces cerevisiae"
 /strain="AB972 - trp1 r(0) (S288C background)"
 /db_xref="taxon:4932"
 /clone_lib="mtn-3xHA/lacZ Insertion Library, strain AB972"
 /lab_host="E. coli"
 /note="Vector: pHS6-Sal; A yeast genomic DNA library was
 prepared in pHS6-Sal; genomic DNA was size-fractionated
 (DNA of roughly 2-3 kb in length) prior to cloning. This
 library was subsequently mutagenized with a mtn-3xHA/lacZ
 minitransposon containing lacZ, URA3, and tet resistance."

BASE COUNT 155 a 117 c 114 g 151 t

ORIGIN

Query Match 11.2%; Score 125; DB 17; Length 537;
 Best Local Similarity 64.4%; Pred. No. 3, 4e-22;
 Matches 204; Conservative 0; Mismatches 110; Indels 3; Gaps 1;

QY 4 TCAGTTCCAACTCAAAAGCCGTCATCATTTGAAGTGACAAAGCTGTGTTAAACA 63
 |||||
 Db 415 TCGATTCCAGAAACATGAAAGCCGTTGCTATTTAAATGCAAGCTGTATGTAACAG 356
 |||||

QY 64 GATGTCAGTTCCAGAAATTAAGAGGGTACAGCCTTGTGAAGGTTGAGGCTGTGCT 123
 |||||
 Db 355 GACATTCCTCAATTCCTGAATTAGAGAGGATTTGTTCTTAATTAAGACTGTGCGCGTGC 296
 |||||

QY 124 GGTAAACCACTGATTTGGAAGCATATGCTTATTAAGTTGGTCCAGAAAGTTCAATTCTA 183
 |||||
 Db 295 GGTAAACCACTGATTTGGAAGCATATGCTTATTAAGTTGGTCCAGAAAGTTCTTA 236
 |||||

QY 184 GGATGTGACATTTGCTGTGTAAGTTGTCAAACTTGACCAATGCTAGTAGTG--ACTTG 240
 |||||
 Db 235 GGCTGTATGACAGCGGCCAAATGTAAGTTGGGCCCAATGTTGATGCTGACAGCTTT 176
 |||||

QY 241 AAGTTGAGATACCGGTTTCGTTTGTTCACGGTCTTCCCAACAGATCTTAAT 300
 |||||
 Db 175 GCCATTGGTATTACATTATAGGGGTTATTCAGGCTCTCAGTAGGTTCCCTCAAC 116
 |||||

QY 301 GGTCATTTCCTGTAATA 317
 |||||
 Db 115 GGTGCTTTCCTGAGTA 99

RESULT 11
 AO502797/c 529 bp DNA linear GSS 29-APR-1999
 LOCUS V75D10 mtn-3xHA/lacZ Insertion library Saccharomyces cerevisiae
 DEFINITION genomic 5', DNA sequence.
 ACCESSION AO502797
 VERSION AO502797.1 GI:4705343
 KEYWORDS GSS.
 SOURCE baker's yeast.
 ORGANISM Saccharomyces cerevisiae
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 RENCE 1 (bases 1 to 529)
 THORS Ros-Macdonald, P., Roemer, T., Coelho, P. S. R., Agarwal, S., Kumar, A.,
 deStranges, S. A., Cheung, K.-H., Sheehan, A., Symoniatidis, D., Jansen, R.,
 Umansky, L., Heidtman, M., Nelson, K., Iwasaki, H., Kanada, D., Lugo, R.,
 Hager, K., Miller, P., Roeder, G. S. and Snyder, M., Kanada, D., Lugo, R.,
 Large-Scale Analysis of the Yeast Genome by Transposon Tagging and
 Gene Disruption
 Unpublished (1999)
 TITLE Contact: Kumar A
 JOURNAL Michael Snyder, Dept. of Mol. Cell. and Dev. Biology
 COMMENT Yale University
 P.O. Box 208103, New Haven, CT 06520-8103, USA
 Tel: 203 432 9949
 Fax: 203 432 6161
 Email: anuj.kumar@yale.edu
 te of mtn-3xHA/lacZ insertion.
 Seq primer: GGCTTCTTCTTTGGAAAGTAC
 Class: transposon-tagged.
 Location/Qualifiers
 FEATURES
 source 1. 529
 /organism="Saccharomyces cerevisiae"
 /db_xref="taxon:4932"
 /clone_lib="mtn-3xHA/lacZ Insertion Library"
 /lab_host="E. coli"
 /note="Vector: pHS6-Sal; A yeast genomic DNA library
 (lacking mitochondrial DNA) was prepared in pHS6-Sal;
 genomic DNA was size-fractionated (DNA of roughly 2-3 kb
 in length) prior to cloning. This library was
 subsequently mutagenized with a mtn-3xHA/lacZ
 minitransposon containing lacZ, URA3, and tet resistance."

BASE COUNT 152 a 117 c 112 g 147 t

ORIGIN

Query Match 11.1%; Score 124; DB 17; Length 529;
 Best Local Similarity 64.0%; Pred. No. 6, 2e-22;
 Matches 203; Conservative 0; Mismatches 111; Indels 3; Gaps 1;

QY 4 TCAGTTCCAACTCAAAAGCCGTCATCATTTGAAGTGACAAAGCTGTGTTAAACA 63
 |||||
 Db 415 TCGATTCCAGAAACATGAAAGCCGTTGCTATTTAAATGCAAGCTGTATGTAACAG 356
 |||||

QY 64 GATGTCAGTTCCAGAAATTAAGAGGGTACAGCCTTGTGAAGGTTGAGGCTGTGCT 123
 |||||
 Db 355 GACATTCCTCAATTCCTGAATTAGAGAGGATTTGTTCTTAATTAAGACTGTGCGCGTGC 296
 |||||

QY 124 GGTAAACCACTGATTTGGAAGCATATGCTTATTAAGTTGGTCCAGAAAGTTCAATTCTA 183
 |||||
 Db 295 GGTAAACCACTGATTTGGAAGCATATGCTTATTAAGTTGGTCCAGAAAGTTCTTA 236
 |||||

QY 184 GGATGTGACATTTGCTGTGTAAGTTGTCAAACTTGACCAATGCTAGTAGTG--ACTTG 240
 |||||
 Db 235 GGCTGTATGACAGCGGCCAAATGTAAGTTGGGCCCAATGTTGATGCTGACAGCTTT 176
 |||||

[illegible]

		/note= "similar to Saccharomyces cerevisiae ORF YCR102c [
		similar to C carbonum toxo gene]
		1 putative framehift(s)
		similar to Saccharomyces cerevisiae ORF YLR640c [
		similar to C.carbonum toxo protein]"
		/evidence=not experimental]
BASE COUNT	252 a 219 c 200 g 271 t	4 others
ORIGIN		
Query Match	10.7%, Score 119; DB 17; Length 946;	
Best Local Similarity	59.8%; Pred. No. 1.5e-20;	
Matches 253; Conservative	1; Mismatches 161; Indels 8; Gaps 3	
Qy	32 TCATTGAAGGTGACAAAGCTTTTGTAAAAACAATGTCACATTCCCAATAAAGAAGAG	91
Db	524 TTATTGAAGGGGAAGAAGCTGTCAATCAAAGAGGGTGCCCATGCCAATTAGAGATG	583
Qy	92 GTACAGCCTTGSTGAAGGTGAGGCCTGTGGTGCGTAACCCAATGATTGGAACATATTG	151
Db	584 GTTTCGTTTCGTATCAAAAACGCTGCTGTGCCGCGTAATCCACTGATTTGGGACATATTG	643
Db	152 CTATTAAGATTGTCACAGAAGTTCAATTCATGAGATGTGACATTTCTGTGACAGTTGTCA	211
Qy	644 ACTCAAGAATATGAGCCCTCAAGGTTCTACTCGGGGGGAGATGCTACAGGTCATATGTTA	703
Db	212 AACTTGAACCAAT---GCTAGTACTGACTGGAAGTTGGAGATACCAGTTTCGTTTTG	268
Qy	704 AACGTGGTCCGGGAGTGCATGACAAACCCCTTTGCCATTGGAGATTAACAGGGTTTA	763
Db	269 TTCACGGTGTCTCCCAAACAGATCTTAAAAATGGTGCAFTTGGCTGAATATGCAAGGTTT	328
Qy	764 TTAGCGTGCTTCCTCGTAGATTTCCCVCYCAATGAGTCCCTTCCCTGAATATTTGCAATT	823
Db	329 ATCCACCTTTGTTTAAACA---GAGTAACTTAACCTCACTCACTGCTGATGAAA--TTTC	383
Qy	824 CATCGAGAGATGCCAPACAAATCATCCAAACGATATTAATGTGTGTGAAGATAGCTAC	883
Db	384 TGAAGCCCTGTGAAGAACTTCGAATCTGTGCATCATTGCCAGTTTGGTTGACAATGCG	443
Qy	884 CTGAAGTCCCCGTAAAGCTTTTGAAGAGGCGCTGTCACTTTCGSGTGAATGACMACTGC	943
Db	444 TGG 446	
Qy	944 CGG 946	
RESULT 13		
A0873736/c	542 bp DNA linear	GSS 08-NOV-1999
LOCUS	V74E1 mtfn-3xHA/lacZ Insertion Library, strain AB972	Saccharomyces
DEFINITION	Cerevisiae genomic 5', DNA sequence.	
ACCESSION	A0873736	
VERSION	A0873736.1 GI:5285980	
KEYWORDS	GSS.	
SOURCE	Baker's yeast.	
ORGANISM	Saccharomyces cerevisiae	
REFERENCE	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;	
AUTHORS	Saccharomycetales; Saccharomycetaceae; Saccharomyces.	
TITLE	1 (bases 1 to 542)	
JOURNAL COMMENT	Ross-Macdonald,P., Roemer,T., Coelho,P.S.R., Agarwal,S., Kumar,A., deSteges,S.A., Cheng,K.-H., Sheehan,A., Symoniatidis,D., Jansen,R. Umansky,L., Heidtman,M., Nelson,K., Iwasaki,H., Kanada,D., Iugo,R. Hager,K., Miller,P., Roeder,G.S. and Snyder,M. Large-Scale Analysis of the Yeast Genome by Transposon Tagging and Gene Disruption Unpublished (1999) Contact: Kumar A Michael Snyder, Dept. of Mol. Cell. and Dev. Biology Yale University P.O. Box 208103, New Haven, CT 06520-8103, USA Tel: 203 432 9949 Fax: 203 432 6161 Email: amj.kumar@yale.edu	

te of mtN-3xHA/lacZ insertion.
Seq primer: GGCCCTCTCTCTTGGAAATAC
Class: transposon-tagged.

FEATURES

source

Location/Qualifiers

1. 542
/organism="Saccharomyces cerevisiae"
/strain="AB972 - trp1 r(0) (\$288C background)"
/db xref="taxon:4932"
/clone_1lb="mtN-3xHA/lacZ Insertion Library, strain AB972"
/lab host="E. coli"
/note="Vector: pHS6-Sal; A yeast genomic DNA library was prepared in pHS6-Sal; genomic DNA was size-fractionated (DNA of roughly 2-3 kb in length) prior to cloning. This library was subsequently mutagenized with a mtN-3xHA/lacZ minitransposon containing lacZ, URA3, and tet resistance."

BASE COUNT 152 a 136 c 112 g 142 t

ORIGIN

Query Match 10.6%; Score 117.8; DB 17; Length 542;
Best Local Similarity 63.3%; Pred. No. 2,7e-20;
Matches 198; Conservative 0; Mismatches 112; Indels 3; Gaps 1;

QY 8 TTCCAACTCAAAAAAGCCGTATCTTGAAGTGCACAAAGCTGTGTTAAACAGATG 67
DB 409 TTCCAGAAACCATGAAAGCCGTGTCTTGAAGAGCGAGCTGTAGACAGACACA 350
QY 68 TCTCAGTTCAGAAATTAAGAGGGTACACCTTGTGAGAGTGAAGCTGTGTTGTA 127
DB 349 TTCCAAATTCCTGCATTGAAGAGAGATTTGTTCTAATTGAGCTGCGCGTTG 290
QY 128 ACCCACTGATGGAGCATATGCTTAAAGATGTCAGAAAGTTCAATTTAGAT 187
DB 289 ACCATACCGATGGAAACATATGATTTGAGATGTCCTCAAGGTCCTCTTAAGGCT 230
QY 188 GTACATGCTGTGACAGTTGTCAAACTTGACCAATGCTAGTACG---ACTTGAAG 244
DB 229 GTATGACAGCCGGCCAAATGTAAAGTTGAGCCCAAAATGTTAGTCAAGCTTGCCA 170
QY 245 TTGGAGATACCGGTTTGGTTTGTTCACCGTGTCTCCAAACAGATCTTAAATG 304
DB 169 TTGGTATTTACATTTAAGGGGTTATTTACCGTGTTCAGTGAAGTTCCCTCAACG 110
QY 305 CATTGCTGAATA 317
DB 109 CATTGCTGAGTA 97

LOCUS T14 337 bp mRNA linear EST 11-JAN-1995
DEFINITION EST104002 S. cerevisiae strain X2180-1A Saccharomyces cerevisiae
VERSION T18505
KEYWORDS T18505.1 GI:622322
SOURCE EST.
ORGANISM Baker's yeast.
REFERENCE Saccharomyces cerevisiae
AUTHORS Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
COMMENT Saccharomycetales; Saccharomycetaceae; Saccharomyces.
1 (bases 1 to 337)
Weinstock, K.
Saccharomyces cerevisiae CDNAS
Unpublished (1995)
Contact: Weinstock, K. and Venter, J.C.
The Institute for Genomic Research
932 Clapper Rd, Gaithersburg, MD 20878
Tel: 3018699056
Fax: 3018699423
Email: tdbinfo@cd.igf.org
For clone availability please contact the TIGR Database
(tdbinfo@cd.igf.org)
Seq primer: M13 Reverse.

FEATURES

source

Location/Qualifiers

1. 337
/organism="Saccharomyces cerevisiae"
/strain="X2180-1A"
/db xref="taxon:4932"
/clone_1lb="S. cerevisiae strain X2180-1A"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 83 a 75 c 77 g 97 t 5 others

ORIGIN

Query Match 10.4%; Score 115.4; DB 14; Length 337;
Best Local Similarity 63.8%; Pred. No. 1e-19;
Matches 203; Conservative 0; Mismatches 111; Indels 4; Gaps 2;

QY 4 TCAGTTCACCACTCAAAAAAGCCGTATCTTGAAGTGCACAAAGCTGTGTTAAACA 63
DB 2 TCGATTCCAGAAACCATGAAAGCCGTTCATTGAAATGCAAGGCTGTATCAACG 61
QY 64 GATGCTCAGTTCAGAAATTAAGAGGTACAGCCCTGTGAGTGAAGCTGTGCT 123
DB 62 GNCATTCCAATTCCTGAATTAAGAAAGGATTTGTTAATTAAGACTGTCCCGTTGCC 121
QY 124 GGTAAACCACTGATTTGAAGCATATGCTTAAAGATTTGTCGCAAGGTTCAATTC 183
DB 122 GGTAAACCACTGATTTGAAGCATATGCTTAAAGATTTGTCGCAAGGTTCCCTCTTA 181
QY 184 GATGATGATGCTGTGACAGTTGTCAAACTTGACCAATGCTAGTACTG---ACTTG 240
DB 182 GCGTGTGATGACAGCCGGCCAAATGTAAAGTTGAGGCCCAATGTTNATGCTCAGCTTT 241
QY 241 AAGTTGATGATACCGGTTTGTGTTTGTTCACCGTGTCTCCAAACAGAT-CTTAAAA 299
DB 242 GCCATTTGGATTAATTAATTAATGAGGTTTATCAAGGTCCTTCAAGTGAAGTTCCCTCAA 301
QY 300 TGTGATTTGCTGAATA 317
DB 302 CCGTGTGCTGAGTA 319

RESULT 15

AZ926593

LOCUS AZ926593 540 bp DNA linear GSS 01-APR-2001
DEFINITION 476.d1809b12.sl Saccharomyces castellii NRRL Y-12630 Saccharomyces
castellii genomic clone 476.d1809b12.sl, DNA sequence.

ACCESSION AZ926593

VERSION AZ926593.1 GI:13497495

KEYWORDS GSS.

SOURCE Saccharomyces castellii.

ORGANISM Saccharomyces castellii.

REFERENCE Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

COMMENT Saccharomycetales; Saccharomycetaceae; Saccharomyces.

1 (bases 1 to 540)

Clifton, P.F., Hillier, L.W., Fulton, L., Graves, T., Miner, T., Gish

W.R., Waterston, R.H. and Johnston, M.

Surveying Saccharomyces genomes to identify functional elements by

comparative DNA sequence analysis

Unpublished (2001)

Department of Genetics

Washington University Medical School

Box 8232, 4566 Scott Ave., St. Louis, MO 63110, USA

Tel: 314 362 2735

Fax: 314 362 7855

Email: mj@genetics.wustl.edu

Class: random plasmid subclone.

Location/Qualifiers

1. 540

/organism="Saccharomyces castellii"

/strain="NRRL Y-12630 (CBS 4309)"

/db xref="taxon:27288"

/clone="476.d1809b12.sl"

/clone_1lb="Saccharomyces castellii NRRL Y-12630"

THIS PAGE BLANK (USPTO)